Two Loops, Three Loops, or Four Loops: Pedagogic Issues in Explaining Basic Epidemic Dynamics

James M. Lyneis and Debra A. Lyneis
Worcester Polytechnic Institute, jmlyneis@wpi.edu
Creative Learning Exchange, lyneisd@clexchange.org
P.O. Box 121, Weston, VT 05161, USA

ABSTRACT

How many feedback loops, and of what type, control the behavior of an epidemic? This seemingly simple question arose on trying to relate the behavior of the epidemic model widely used in K-12 system dynamics to its three-loop feedback structure. A search of the literature discovered two-, three-, and four-loop versions of the basic epidemic model in introductory system dynamics materials. How can the same behavior be explained with such different feedback structures? Can they all be right? This paper analyzes the three basic model structures and discusses implications for system dynamics pedagogy. We conclude that either the two- or four-loop versions of the basic epidemic model are acceptable representations, with the two-loop version recommended for beginners; the three-loop version of the system is never correct. In addition, we suggest that the development of incorrect representations such as the three-loop epidemic model can be avoided if standard system dynamics modeling practice is followed. At a macro level, standard practice dictates first formulating a dynamic hypothesis to explain the observed behavior (rather than building models by "hooking stocks and flows together"); at a micro level, standard practice suggests avoiding multiple algebraic expressions within variables.

Introduction

How many feedback loops, and of what type, control the behavior of an epidemic? This seemingly simple question arose on trying to relate the behavior of the epidemic model widely used in K-12 system dynamics to its three-loop feedback structure. A search of the literature discovered two-, three-, and four-loop versions of the basic epidemic model in introductory system dynamics materials. How can the same behavior be explained with such different feedback structures? Can they all be right?

Does it matter how many feedback loops produce the behavior? Why not just simulate the model, see what happens, and conclude that "the model" produced the behavior? Good system dynamics practice demands that we relate feedback structure to behavior in order to explain ***why*** a simulated behavior occurs. Such an explanation serves three purposes. First, understanding how structure creates behavior helps us to develop our intuition about how feedback systems behave and transfer that intuition to other situations. Second, it helps us to more effectively design policies to change that behavior. And third, it helps convince us (or decision-makers) to do something. People rarely take action because "the model says so." They need to have some intuitive sense as to why a change in behavior would be helpful, and the feedback loop explanation provides that sense.

In trying to understand epidemic behavior, we looked at several introductory descriptions of the epidemic model, and found ***three different versions of exactly the same equation structure, all producing identical behavior:***

- A two-loop version (Sterman, 2000)
- A three-loop version from Road Maps (Glass-Husain, 1991)
- A four-loop version from the WPI introductory system dynamics course (Hines and Lyneis, 2005)

In this paper, we will examine and compare the three versions and discuss implications for system dynamics pedagogy and the goal of relating structure to behavior.

Basic Epidemic Models – Constant Total Population

The Epidemic Behavior Mode

The behavior of an epidemic in which the entire population becomes infected is illustrated in Figure 1. The behavior of the infected population looks similar to the S-shaped "limits to growth" archetypal behavior. Therefore, it seems that the structure which creates that behavior might include a positive feedback loop (which is initially dominant) creating growth in infections, and a negative feedback loop (dominant later in the behavior) which reduces infections as the healthy population declines. Is that the case? And if so, what is the goal (implicit or explicit) of the negative feedback loop?

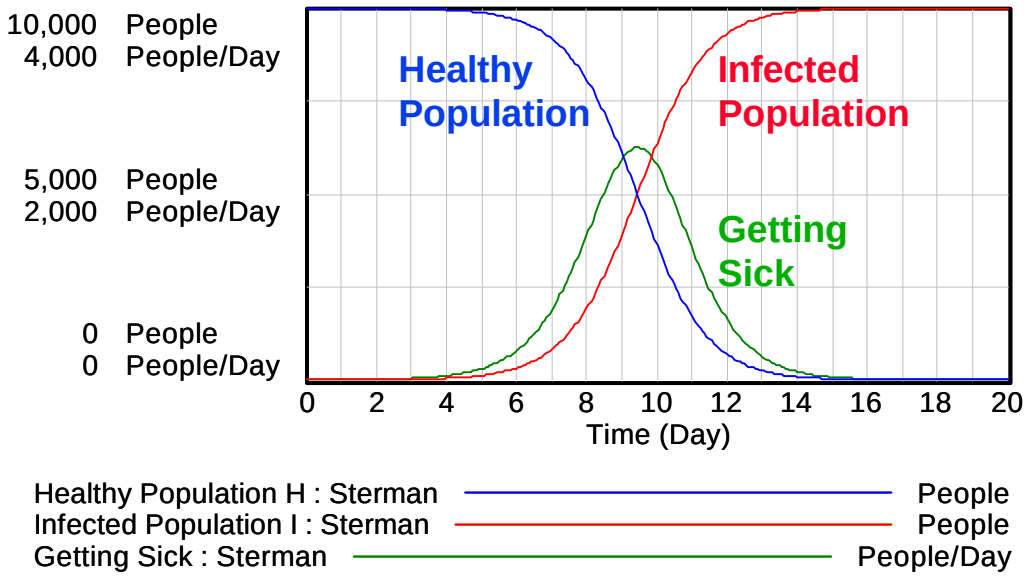


Figure 1: Epidemic Behavior Mode

The Two-Loop Model

The basic two-loop structure, illustrated in Figure 2, is adapted from Sterman's SI (Susceptible and Infectious) model. (Sterman, 301) Key equations are noted in bold italics and some of the variable names are shortened to reduce clutter (we use Healthy and Infected which are more common with K-12 students). Infected people make contacts with others, as determined by the Contact Rate *c* (in people/person/day). Some of these contacts are with healthy people. Assuming equal mixing, the fraction of contacts with healthy people is simply equal to the fraction of healthy people in the population. So,

Healthy Contacts with Infected People = Contacts by Infected People * Fraction Healthy

Getting sick = Healthy Contacts with Infected People * Infectivity of the disease.

A critical assumption is that the Total Population remains constant (we will relax this assumption later). The initial values for Healthy Population and Infected Population are set to sum to Total Population. We further assume that once people become infected they remain infected; for now, there are no recoveries or deaths. There are also no vaccines or quarantines.

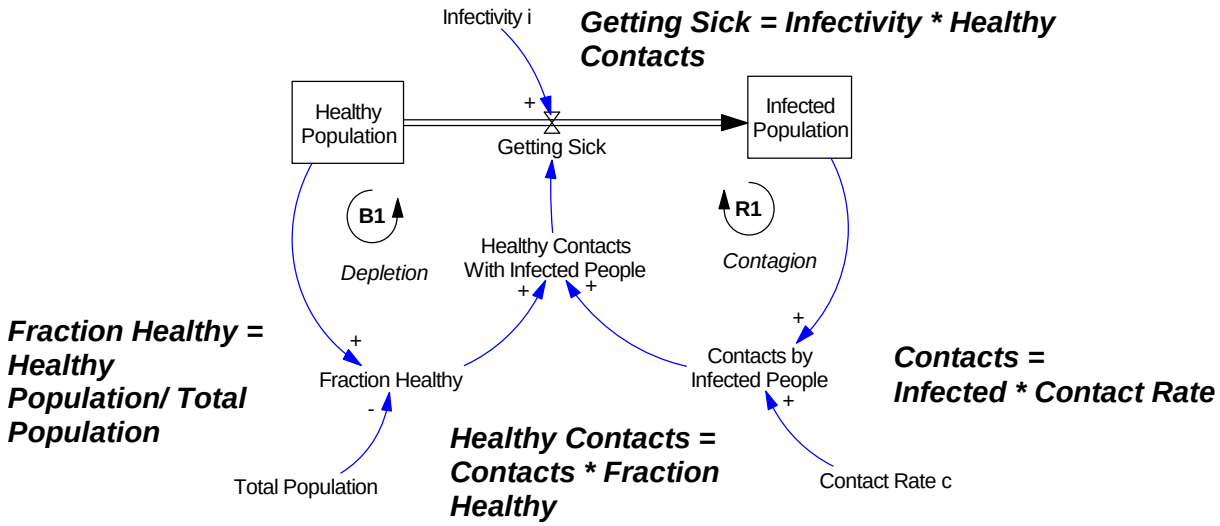


Figure 2: Two-Loop Model (adapted from Sterman)

In this model, there is one positive "contagion" loop and one negative "depletion" loop. As we will see, these two feedback loops are in fact the *only* two loops which are dynamically active in this simple epidemic model. Together they are necessary and sufficient to produce the reference behavior: As the number of infected people increases, their contacts increase and they make more contacts with healthy people. As a result, more people get sick, further increasing the infected population; this positive "contagion" loop dominates the behavior of Infected People in the early exponential growth phase of the epidemic shown in Figure 1.

However, as the healthy population declines, the fraction of the population that is healthy decreases, so the number of healthy contacts with infected people decreases. As a result, the increase in Getting Sick begins to slow and eventually peaks, and the growth in the Infected Population slows. During the period from about Day 9 to Day 11 in Figure 1, both feedback

loops have a significant impact on behavior (and in fact, the dominant loop is shifting from the contagion loop to the depletion loop). However, because Getting Sick is still positive, Healthy Population declines, which further reduces the Fraction Healthy and Getting Sick. This negative "depletion" loop dominates Infected behavior in the asymptotic approach phase late in the epidemic shown in Figure 1. Eventually, the system reaches equilibrium when everyone is infected.

The structure shown in Figure 2 is more complicated than the structure in Sterman's book (Sterman, 301). The simpler original structure is shown in Figure 3. We have included more "auxiliary" variables in our feedback loops in order to give a more "operational" view of how an epidemic works. We have explicitly identified the concept of Contacts by Infected People, the fact that some of these contacts are with Healthy People, and that these latter contacts in turn might lead to Getting Sick. When broken down, the flow equation for Getting Sick in Figure 2 is the same as the equation for Figure 3:

Getting Sick = c * i * Infected Pop * Healthy Pop/Total Pop

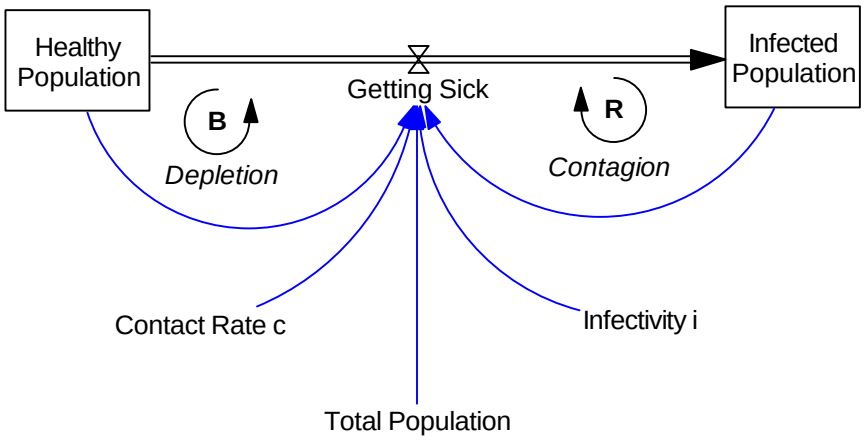


Figure 3: Original Sterman Two-Loop Model

In one sense, the simpler structure shown in Figure 3 is a more "mathematical" view of the system, whereas Figure 2 is a more operational description. Figure 3 collapses the structure to its purest feedback structure. Unfortunately, there are alternative ways of operationalizing the basic epidemic concept, and, as we will show, these alternatives seemingly introduce additional feedback loops.

The Three Loop Model

The three-loop structure, shown in Figure 4, results from an operationalization which recognizes that the healthy and infected populations equal the total population (although this violates a standard system dynamics modeling practice of avoiding multiple algebraic expressions in equations). This model was presented in a Road Maps paper by William Glass-Husain in 1991 as an undergraduate in the MIT System Dynamics in Education Project (Glass-Husain, 54). It has been widely used in K-12 system dynamics lessons since then. Glass-Husain developed the

Epidemic Game so that students could experience the spread of an infection as the spread of a secret handshake through the class of students. Students then graphed the spread of their infection and used that as their reference behavior mode to build the model. In 2004, Quaden, Tickotsky and D. Lyneis adapted a simplified version of the game for younger students; a recent attempt to write up a loop analysis of the game and model led to the current discussion.

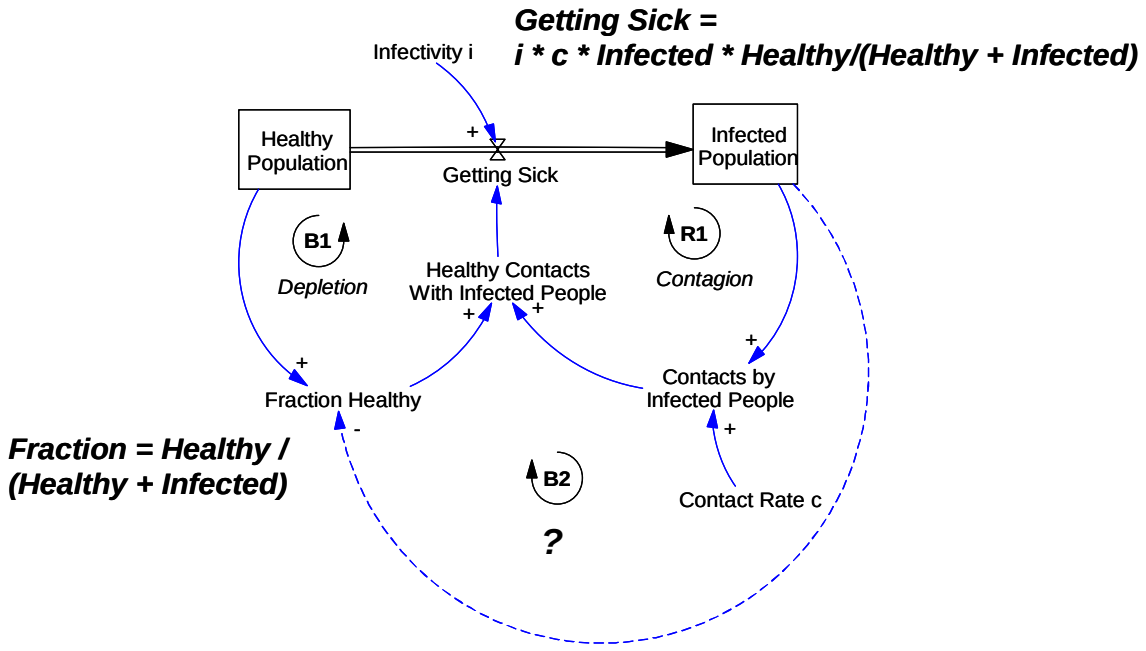


Figure 4: Three-Loop Model (adapted from Road Maps)

The Road Maps paper explained the dynamic hypothesis with a positive contagion and negative depletion loop and built a simple model of those two loops (B1 and R1 above). However, in the Road Maps model, Total Population is not a parameter of the model and does not appear explicitly in the model, as it does in the previous two-loop models. Instead, the equation for Fraction Healthy sums the Healthy and Infected populations to get the equivalent of Total Population for the denominator. In order to get the infected population into the denominator of the fraction equation, this model adds a third connection denoted by a dashed line in Figure 4. Adding this link seems to make sense on the surface, but a closer look raises many interesting questions.

This new feedback loop B2 appears to be a negative feedback loop: If Infected Population increases, Fraction Healthy decreases, thereby decreasing Healthy Contacts With Infected People and, in turn, Getting Sick. This causes the Infected Population to decrease (more accurately, to be lower than it otherwise would be). So the initial increase in Infected Population works its way around the loop to cause a decrease (or relative lowering) of Infected Population.

How does this third feedback loop affect the S-shaped behavior of the epidemic? Since it is a balancing loop and produces a slowing in the growth of infected people, it would seem to contribute to the asymptotic behavior later in the epidemic. One might conclude that the "depletion" loop and the "?" loop together produce the asymptotic behavior. But do they?

Before answering that question, note that it is possible to operationalize an alternate version of the three-loop model which starts with contacts from the Healthy Population and computes the fraction of those contacts with infected people as shown in Figure 5.[1] When collapsed, the equation for Getting Sick in Figure 5 is equal to the equation for Getting Sick in Figure 4.

Equation Fig. 4: Getting Sick = $c * i *$ Infected * Healthy/(Healthy + Infected)
Equation Fig. 5: Getting Sick = $c * i *$ Healthy * Infected/(Healthy + Infected)

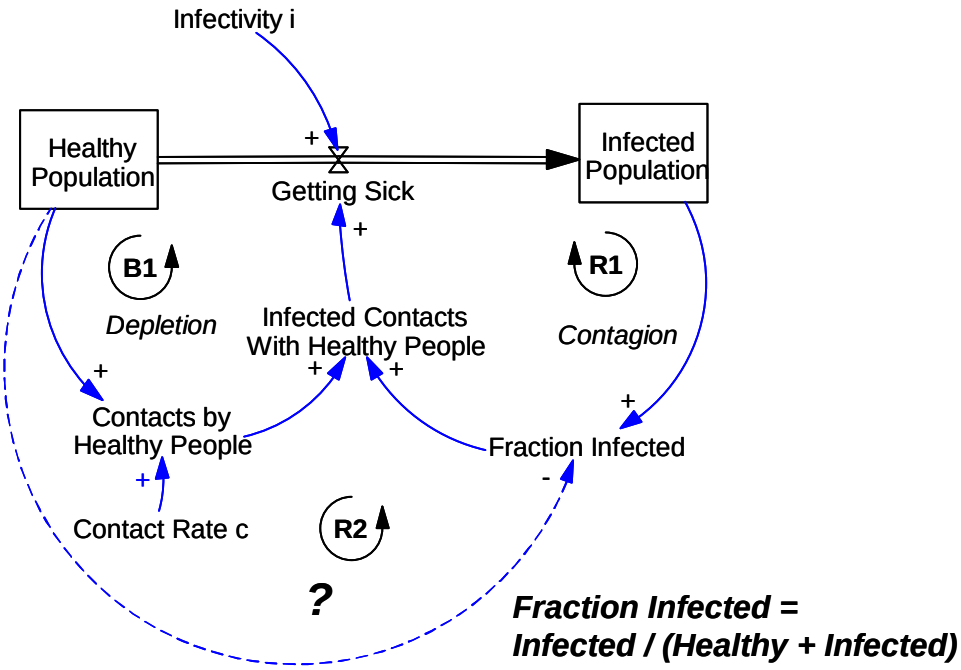


Figure 5: Alternate Version of Three-Loop Model

Yet in this version, a third feedback loop from healthy population is added, but now it is a *re-enforcing* loop! A decrease in the Healthy Population increases the Fraction Infected, which in turn increases Healthy Contacts With Infected People and Getting Sick, such that the Healthy

[1] We could also draw a similar alternate version of the two-loop model with the contacts coming from healthy people. The equations are identical and produce identical behavior. We opt for the contacts with infected formulation because that seems to best represent the idea of infected people spreading the illness through the population, and describes the S-shaped behavior mode of the infected population. It also reflects students' experience with the game as they notice the infection spreading among them. Nevertheless, both formulations are simplifications of reality – in real life, both healthy people and infected people (unless quarantined or bedridden) are out making contacts.

Population decreases – the initial decrease in Healthy Population works its way around the loop to decrease Healthy Population even more, thereby making it a positive loop.

What does this positive loop contribute to the S-Shaped behavior of the epidemic? Because it is a positive loop, it would seem to contribute to the early exponential growth (caused here by the contagion loop R1 in Figure 5). Does this additional "?" re-enforcing loop work with the contagion loop to cause exponential growth early in the epidemic? Moreover, how is it possible that alternative operationalizations of the same structure have different feedback loops, and produce exactly the same behavior? How can we intuitively explain that behavior?

(While we have not shown simulation output, with the same initial stocks, total population, contact rate, and infectivity, all of the models shown above, and the four-loop versions shown later, produce exactly the same behavior as the reference mode shown in Figure 1.)

The Four Loop Model

Before answering the questions about the alternative three-loop feedback structures and examining the impact of these third feedback loops, we will complete the spectrum and look at the four-loop operationalizations of the model adapted from the WPI course, System Dynamics Foundations: Managing Complexity, co-taught by Jim Hines and Jim Lyneis (see Figures 6 and 7).[2] In the four-loop versions of the epidemic model, Total Population is again introduced as an explicit variable, but is represented as the sum of the two stocks rather than as an input parameter, as in the two-loop model. This formulation follows standard system dynamics practice of avoiding multiple algebraic expressions within equations.[3] This formulation seemingly adds a fourth feedback loop to the three-loop model of Figure 4 – a positive loop from Healthy Population through Total Population, Fraction Healthy, Healthy Contacts with Infected People, Getting Sick, and back to Healthy Population.

Again note that it is possible to develop an alternate version of the four-loop model corresponding to the alternate three-loop model of Figure 5. As shown in Figure 7, this alternate version also adds a fourth feedback loop, this time a *negative* loop from Infected Population through Total Population, Fraction Infected, Infected Contacts With Healthy People, Getting Sick, and back to Infected Population. In both cases, the flow equation is once again the same, producing the same behavior. And, unlike the three-loop models, with either formulation the four-loop version of the model has a consistent number and polarity of feedback loops.

Comparing the four-loop to the two-loop version, the two additional connections are added to make the calculation of the Total Population explicit as the sum of the Healthy and Infected populations. What do these fourth loops add to the dynamics? Do they work with the original positive and negative loops to produce the S-Shaped behavior?

[2] In the WPI course, the actual model depicts the diffusion of a new product rather than an epidemic, but the structure is the same.

[3] Avoiding multiple algebraic expressions, however, often clutters up diagrams. Therefore, modelers are tempted to violate this rule in order to simplify their diagrams. The danger of this is the creation of fictitious feedback loops as in the three-loop versions of the epidemic model.

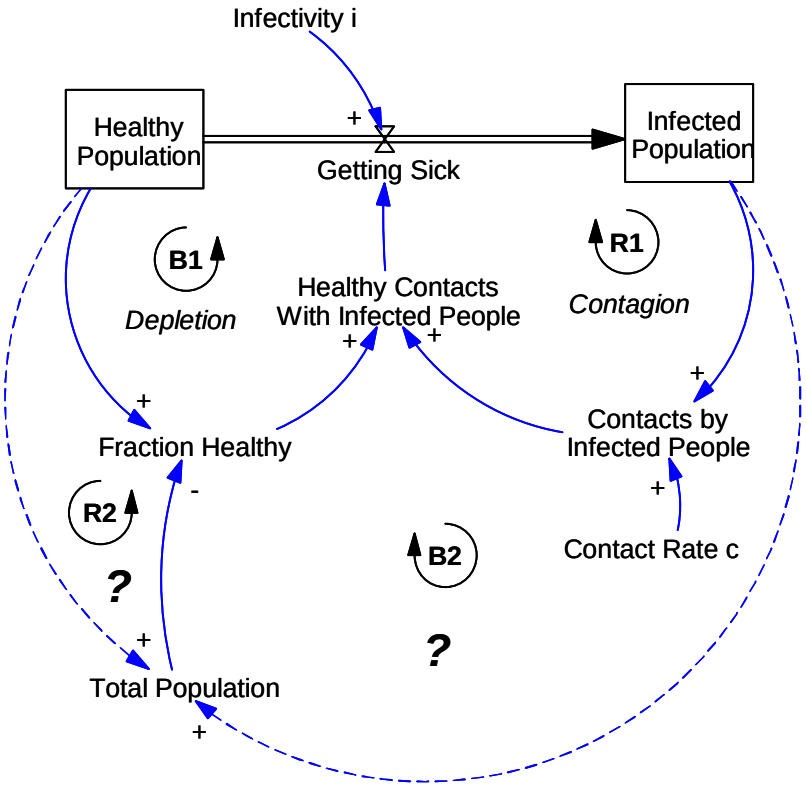


Figure 6: Four-Loop Version (adapted from Hines)

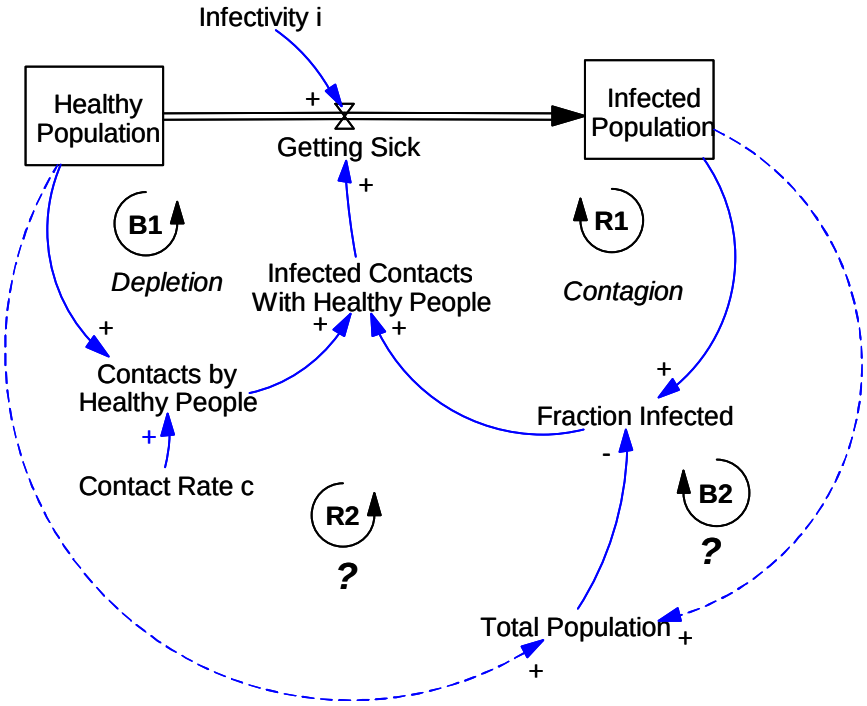


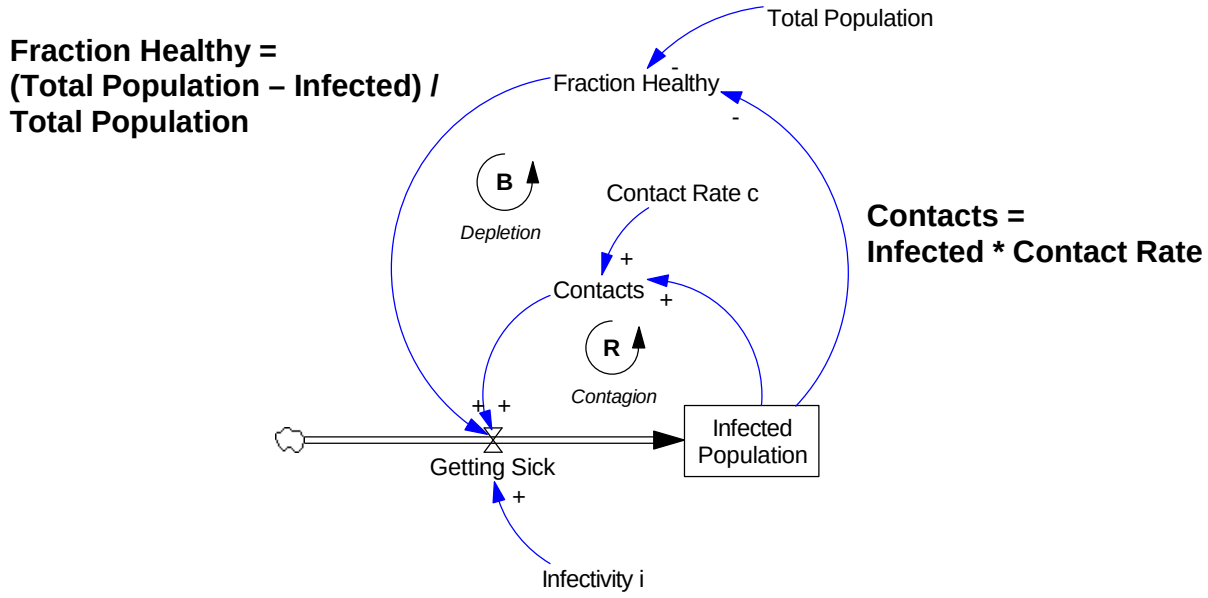
Figure 7: Alternate Version of Four-Loop Model

Analysis

How is it possible to have so many different feedback loop structures for the simple epidemic model? How do the additional feedback loops influence behavior? Finally, is there a correct structure? The first point to notice is that the "extra" feedback loops in the three- and four-loop structures arise because they are used to calculate the total population (implicitly in the three-loop version, and explicitly in the four-loop version). But, because Total Population does not change, these feedback loops are not *dynamically* active. In the four-loop versions of the model, this cancellation is clear. For example, in Figure 7, if the Infected population increases, the Total Population does not increase because the Healthy Population decreases by the same amount so it becomes clear that the "?" balancing and re-enforcing loops exactly cancel each other and therefore are inactive. In the three-loop versions of the model, this cancellation is obscured by the burying of the addition in the denominator. We can say that the third feedback loop is dormant, but this is not visually transparent. For the purposes of explaining behavior, the extra loops are not real feedback loops. They do not contribute to the behavior of the model. Only the original Contagion and Depletion loops determine the S-shaped behavior.

First Order System

This basic point is perhaps easier to see if we recognize that with a constant Total Population, the Healthy and Infected populations are not uniquely determined – one is the Total Population (a constant) minus the other. Therefore, the system is effectively a first-order system and can be structured as shown in Figure 8. Here the fact that there are only two feedback loops becomes evident. It is also becomes clear that the "goal" of the negative loop is to drive the Infected Population to the Total Population, by making Fraction Healthy go to zero – everyone becomes infected.



Getting Sick = Contacts * i * Fraction Healthy
Getting Sick = *c * i * Infected ((Total Pop – Infected)/Total Pop)*

Figure 8: One-Stock Version with Constant Total Population

We might also have come to this first-order representation had we started the model development process by following standard system dynamic practice of first formulating a dynamic hypothesis, or theory, to explain the observed behavior. Doing so would have led us to the causal hypothesis illustrated in Figure 9. The hypothesis represents the re-enforcing contagion loop and the balancing depletion loop. The hypothesis in Figure 9 directly translates into the first-order model shown in Figure 8. In addition, the correspondence of this structure to the "limits to growth" s-shaped growth structure also becomes more apparent. Nevertheless, we do agree that the two-stock version of this model is perhaps a clearer and more operational representation for students, and of course leads more easily to more complex models as discussed in the next section. How we make the connections between the epidemic structure and the limits to growth structure for beginning students remains an open question. The important point here is that if we follow standard system dynamics practice for developing models, we are less likely to introduce spurious feedback loops – had we developed the dynamic hypothesis with two loops, even had we used the two-stock structure, we would have focused on representing the two loops in our hypothesis, and would have immediately questioned the introduction of a third loop.

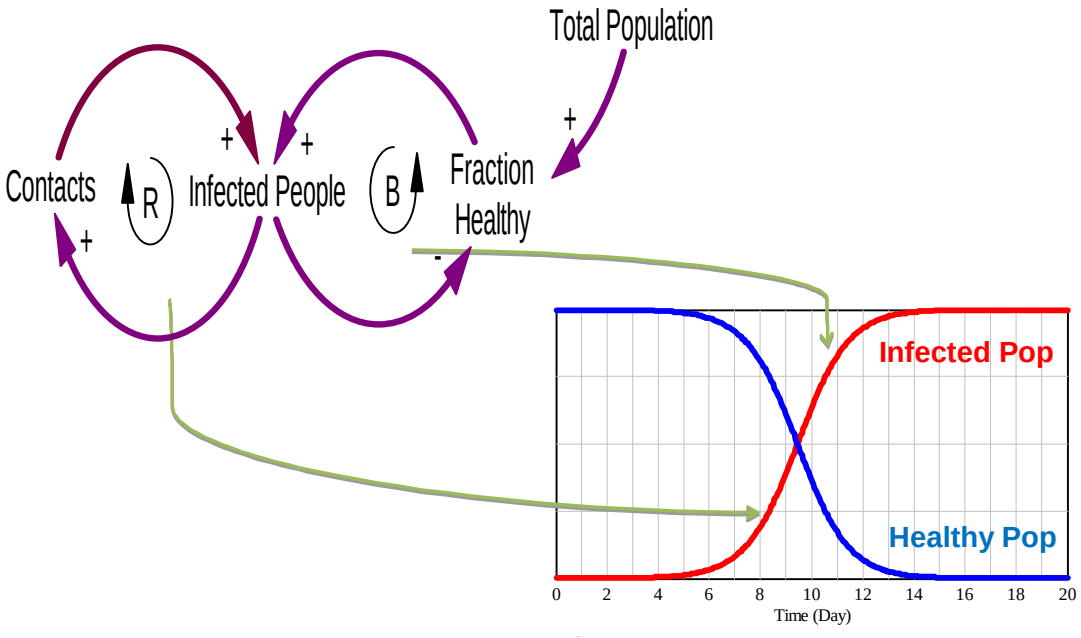


Figure 9: Causal Hypothesis of Epidemic Behavior

Adding Recovery

A second point to notice is that the "extra" third and fourth feedback loops are not dynamic as long as Total Population is constant, even if we add a third stock of recovered population and even a flow back from losing immunity as shown in Figure 10. The basic equation for Getting Sick developed in Figure 2 still applies: Contacts by Infected People is determined only by the Infected Population, and Fraction Healthy is determined only by the Healthy Population divided by the Total Population (a constant).

Getting Sick = c * i * Infected * Healthy/Total Pop

If a third "Recovered" stock is added to the three- or four-loop models above, one additional non-dynamic loop would be added from the Recovered stock to compute the (constant) Total Population, as indicated in Figure 10.

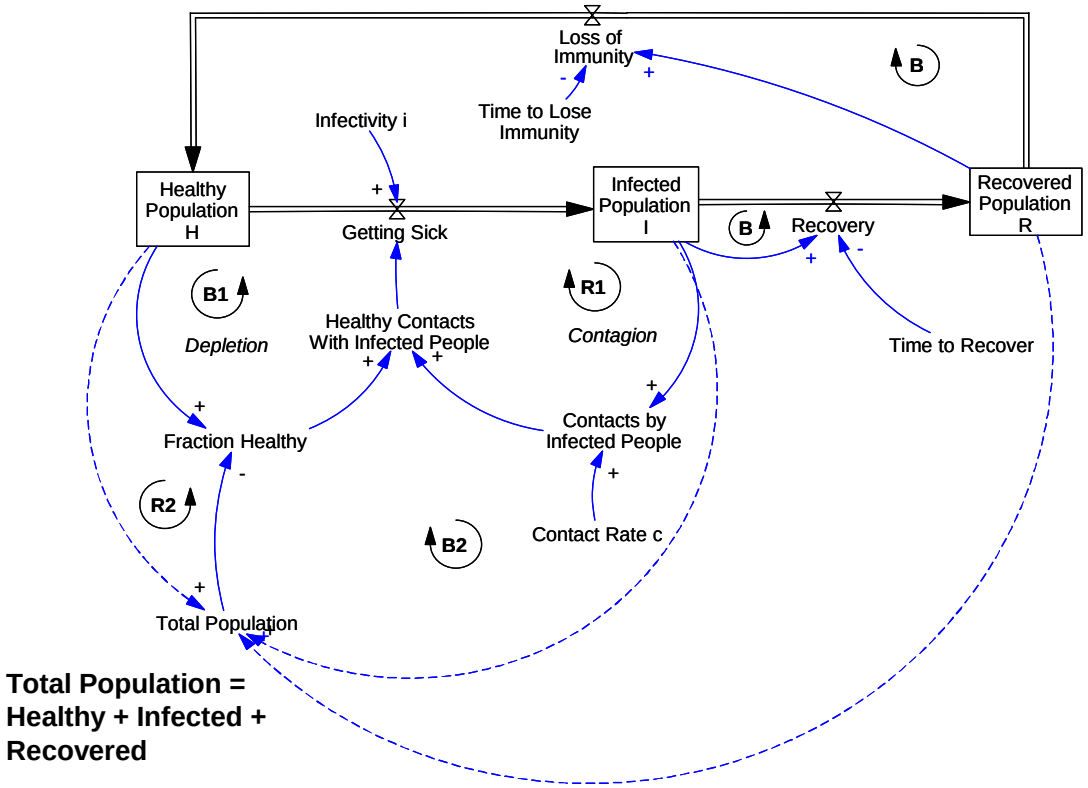


Figure 10: Epidemic Model with Recovery and Loss of Immunity

Results for Constant Population

In summary, if Total Population does not change, then there are only two feedback loops which drive behavior: the Contagion and the Depletion loops of Figure 2. It seems to us that the two-loop structure in Figure 2 is the best representation for beginning students, particularly in K-12, because it focuses attention on structure and behavior without the need to explain why the additional loops created to compute a (constant) total population are not really feedback loops that contribute to the dynamics. We do not find the other possible approaches acceptable: either avoiding any explanation of behavior other than a cursory observation that the population moves from the healthy stock to the infected stock, or explaining the behavior with the two active loops and simply stating that the other loop(s) do not do anything.

If we want to make the point that the sum of the Healthy and Infected stocks sum to Total Population, and that with a constant total population the two loops are inactive, then the four-loop structure with dashed links (Figure 6) is best because it explicitly identifies Total Population as the sum. (As we will discuss later, the added loops are useful in more complex models when advanced modelers understand their purpose.)

In our view, the three-loop structure is misleading and in fact incorrect because it introduces a third loop which is not dynamic, but it is not clear that the loop is always non-dynamic without the fourth loop and explicit identification of Total Population as a concept. Furthermore, the three-loop model obscures an understanding of the direct relationship between feedback structure and behavior.

More Complex Epidemic Models – Changing Total Population

To leave no stone unturned (and to illustrate how a thorough analysis can deepen understanding of a model), what if the total population is changing? For example, what if the Healthy Population grows with births and the Infected Population dies as shown in Figure 11? Growth and deaths add another re-enforcing feedback loop (R3) and another balancing loop (B3), thereby causing the Total Population to change over time. Now, we clearly need feedbacks from the Infected and Healthy Populations to determine the Total Population.

The four-loop version in Figure 11 adds an active positive and an active negative loop; the three-loop version in Figure 12 would add only an active negative link. In this case, which is a better representation? Which is better operationally? Which better helps explain the dynamics?

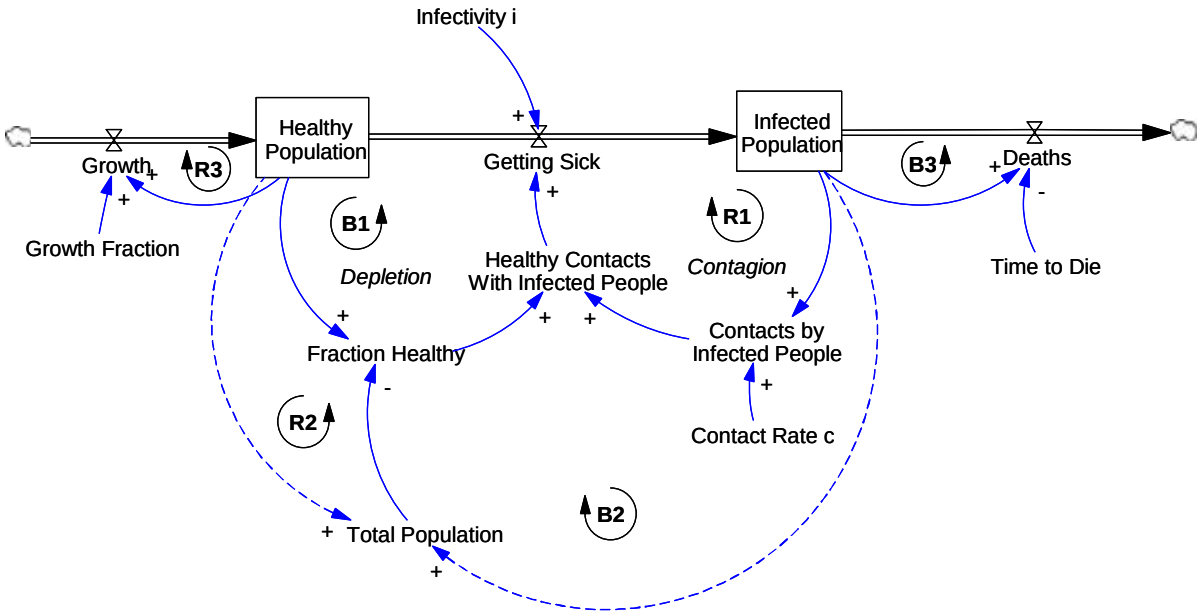


Figure 11: Four-Loop Model with Changing Population

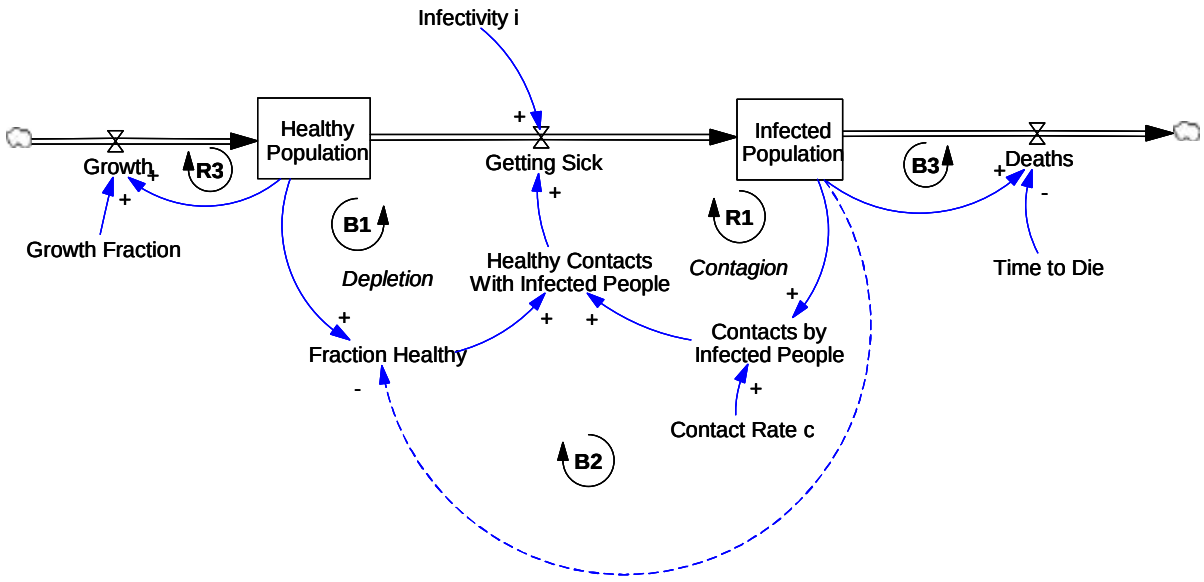


Figure 12: Three-Loop Model with Changing Population

The advantage of the four-loop structure is that, again, it explicitly identifies Total Population as a separate concept. The four-loop version gives a better "operational" picture of the structure. A possible disadvantage of the four-loop structure, however, is that it adds two feedback loops through one variable, "Fraction Healthy." As noted by John Lyneis in a personal communication, showing the link through Total Population from Healthy Population introduces both a negative and a positive feedback loop through Fraction Healthy. Are there really two feedbacks from one variable (Healthy Population) through the same variable (Fraction Healthy)? After all, Fraction Healthy is only one causal effect on getting sick. Can there really be two feedback loops from the same variable for one causal effect? Note that the link from Infected Population to Fraction Healthy adds a second feedback effect from Infected Population to Getting Sick, but here the feedback arises through two different channels: Contacts and Fraction Healthy.

Moreover, the "positive" feedback loop from Healthy Population does not seem to be logical: an increase in Healthy Population increases Total Population, which decreases Fraction Healthy people, thereby decreasing Healthy Contacts with Infected People and Getting Sick, and increasing Healthy Population. How can an initial increase in Healthy Population increase Healthy Population further by Getting Sick? The answer to this last quandary is in the interpretation of feedback loops through stocks and flows (and illustrates the similar problem encountered interpreting loop polarity from causal loop diagrams). A decrease in Getting Sick does not increase Healthy Population, but rather causes Healthy Population to be "greater than it otherwise would be." So, if you work your way around the loop, you see that an increase in Healthy Population, through the positive loop, causes Healthy Population to be greater than it otherwise would be, not to increase.

Nevertheless, the issue of whether or not there can be two separate feedback loops from one variable (Healthy Population) through one causal factor (Fraction Healthy) remains. Clearly

changes in Healthy Population have both a positive and a negative impact on Healthy Fraction, "other things remaining equal." But for these links, other things do not remain equal as Healthy Population must simultaneously drive both links. Should these links therefore be collapsed into one link for the purposes of feedback analysis? And if so, how do we separately, and operationally, show the concept of Total Population?[4]

Perhaps the three-loop structure is more "dynamically" correct? On the other hand, since the two links from Infected Population to Getting Sick also change simultaneously, are there really two feedback effects from Infected Population? Or, for the purposes of feedback analysis, should we collapse the structure to the two feedback loops as shown in Figure 13? Are there other structures which present this dilemma?

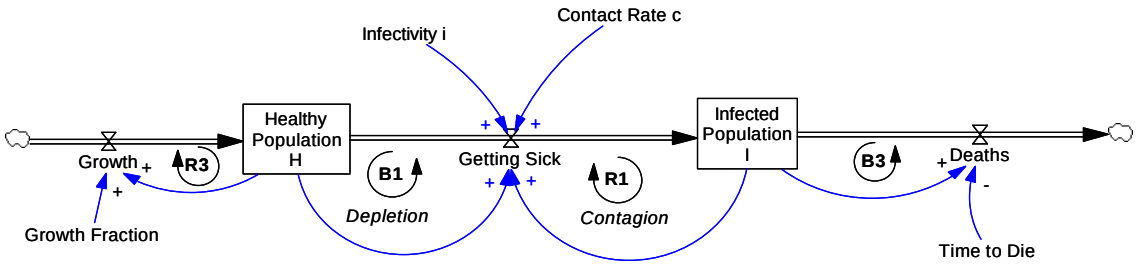


Figure 13: Epidemic Model Collapsed to Simplest Feedback Structure

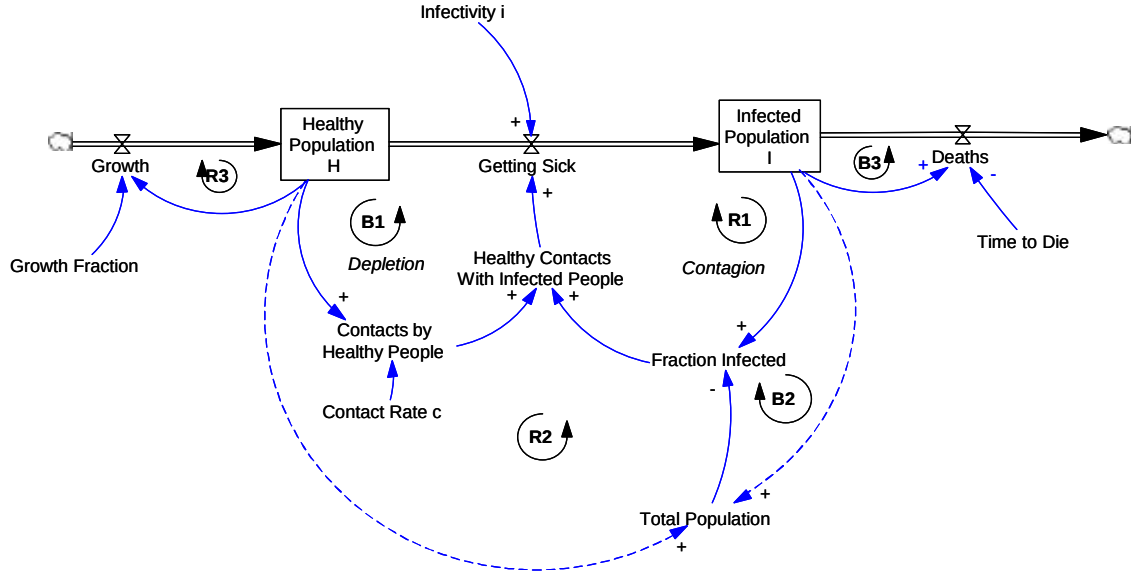


Figure 14: Alternate Four-Loop Model with Changing Population

[4] As a final complication, consider the alternate version of the four-loop model where contacts are driven by Healthy People, and Fraction Infected determines Healthy Contacts With Infected People as shown in Figure 13. With changing population, the "double-effect" occurs from Infected Population through Fraction Infected.

Mohammad Mojtahedzadeh has developed *Digest*, software that calculates the Pathway Participation Matrix to detect which feedback loops in a model are most influential in determining a variable's behavior. (Mojtahedzadeh, *et al*, 2004) In personal correspondence, Mojtahedzadeh reports that according to *Digest* the third loop in the three-loop version remains dormant throughout the simulation. In the four-loop model, the Contagion loop is dominant at first until the Depletion loop takes over, while the extra loops to calculate the Total Population achieve "perfect cancellation" (raising more good questions about whether those loops are "non-dynamic" or "dormant" and the complexity introduced when the cancellation is not perfect.)

Loop Analysis: Changing Population Growth Rates Only

Which structure helps us better explain system behavior? Figure 15 shows the behavior of the epidemic structure with increasing growth fraction, but no deaths (Time to Die is set to 1e9). Growth Fraction ranges from 0.01 per day (approximately 3.65%/year) to 0.1 per day (approximately 35.5% per year). These are high values for population growth, but illustrate the dynamics. The behavior mode initially is a growth in the healthy population, followed by an S-shaped decline; the infected population exhibits S-Shaped behavior, with the saturation level increasing as the growth fraction increases. So the addition of the growth feedback in population causes Healthy Population to increase initially (the higher the growth rate, the higher and longer the increase), but once the Contagion loop begins to take hold the S-shaped dynamics of the epidemic play themselves out, seemingly without much effect from the growth loop R3.

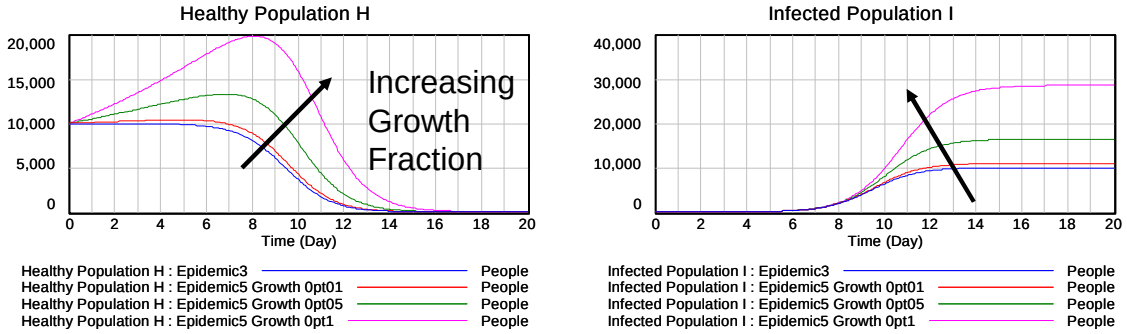


Figure 15: Epidemic Behavior with Increasing Growth Fraction Only

This is perhaps clearer if we examine the behavior of Getting Sick shown in Figure 16. The same basic shape of initial exponential growth, followed by a slowing and then peak, followed by an asymptotic decline occurs regardless of the growth fraction. This pattern of behavior is driven by the Contagion and Depletion loops of the epidemic. The growth loop R3 affects the

ultimate level of the Infected and Total Population, but not the basic contagion behavior mode.[5]

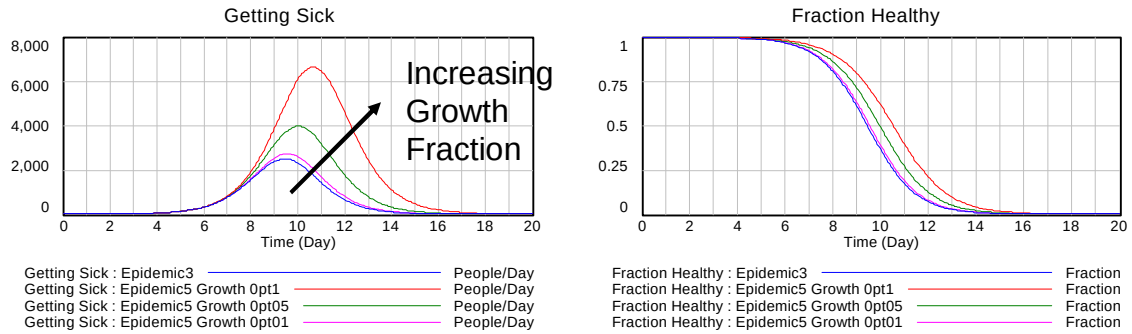


Figure 16: Epidemic Behavior with Increasing Growth Fraction Only

So do the loops R2 and B2 through Total Population in Figure 11 add anything to our explanation of the dynamics? (Note that the blue curve "Epidemic 3" is the behavior with growth fraction of 0.0.) As the infected population grows, the impact of the re-enforcing Contagion loop is clear. At the same time, growth in Infected Population has a depressing effect on Fraction Healthy, but this dynamic is hard to observe in the simulation output. While Infected Population has both a re-enforcing and a balancing effect through Getting Sick, the re-enforcing effect is clearly dominant. The balancing loop B2 does not add anything significant to our understanding of the dynamics. Similarly, the impact of the Depletion loop becomes dominant in the decline phase of the epidemic. This is caused by the reduction in Healthy Population and Fraction Healthy. While this decline in Healthy Population has a balancing and a re-enforcing effect through Fraction Healthy and Getting Sick, the re-enforcing loop R2 does not add anything to our understanding of the dynamics.

Loop Analysis: Changing Death Rates Only

Figures 17 and 18 show the behavior of the epidemic structure with changing Time to Die, but no growth (Growth Fraction is set to 0). In this situation, the behavior mode can be affected by an outflow of deaths from the infected stock (note that the dynamics would be similar if the Deaths flow represented loss of infectivity as well, but the values would differ because contact fractions would change). As Time to Die decreases, the Infected Population increases more slowly. As a result, the epidemic "contagion" dynamics are weakened. If Time to Die is low enough (here 1 day), the epidemic does not occur as the Infected Population dies off before it gets a chance to make contact with the Healthy Population. However, except in this extreme case, even with death (or loss of infectivity) of the infected population, the Contagion and Depletion loops do play themselves out eventually. Again the epidemic dynamics are explainable with the R1 Contagion and B1 Depletion loops; the addition of loops R2 and B2 do not add anything to our understanding of the basic dynamics.

[5] Note that if the Infected population also contributes to growth in the Healthy population, the basic epidemic behavior remains (unless the growth rate is very high), except that the Healthy population never reaches zero, and it, Getting Sick, and Infected population grow. If the growth rate is high enough, the growth loop can dominate and swamp the s-shaped epidemic behavior mode.

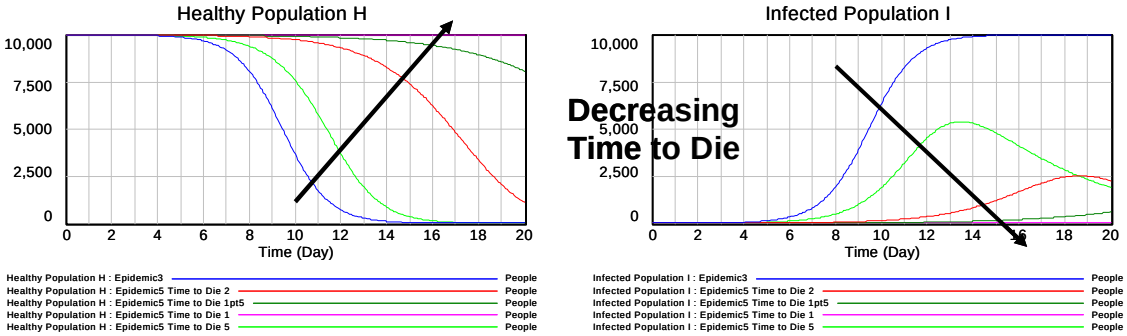


Figure 17: Epidemic Behavior with Increasing Time to Die Only

Figure 18: Epidemic Behavior with Decreasing Time to Die Only

In summary, in the situation where population changes as a result of growth of the healthy population and/or death of the infected population, the four-loop version of the epidemic model (Figure 10) provides a clearer operational view of the system, but the additional re-enforcing loop R2 and balancing loop B2 do not contribute much to the epidemic dynamics, and therefore to our understanding of the causes of these dynamics. The two-loop version of the model (Figure 12) is less operational, but captures the key dynamics in an easily explainable manner. The three-loop version of the model (Figure 11) is less operational than the four-loop version, and does not add to our understanding of the dynamics. There do not seem to be any good arguments for representing an epidemic with the three-loop structure.

Other Analyses

In the previous analyses, we assumed that the total population changed only through growth of the healthy population or deaths in the infected population. In the real world, however, there is more to the story: healthy people can also die and infected people can add to the population. Would these changes to the total population affect the dynamics of the epidemic? In our example, the time frame of the epidemic is only twenty days, so population growth does not make sense (unless people are entering the region), and only deaths by the infected population as a result of the disease would be relevant. For an epidemic of longer duration, such as HIV/AIDS, these other loops would be relevant. However, the key dynamic points would remain – the s-shaped epidemic pattern emerges unless the growth rate is very high, or the time to die very small. In all cases where the epidemic pattern exists, it is caused by the two dominant

feedback loops – the contagion and depletion loops of Figure 2. No other loops are needed to explain that behavior.

However, do we really need to bother with probing more deeply now that we've already looked at the simple epidemic model in so many ways? The purpose of building a model is to build a deep intuition about how feedback systems work – not just to produce a structure that generates the expected behavior. We build that intuition through a series of carefully designed simulation experiments until we understand how the system works under a wide range of conditions. To understand what causes the s-shaped pattern of behavior, we need to see what conditions cause the pattern to persist and what conditions change the shape entirely. This process of analysis provides the reasoning for taking action in this system and an understanding that transfers to other systems.

Conclusion

Are there broader implications for system dynamics practice and pedagogy arising from this interesting little model? A key question is: what is the proper balance between simplifying to focus on feedback loop structures and adding complexity to make a model more operational for clearer connection to real-world phenomena? Experienced system dynamicists typically work on more complex models. These models have many more variables and feedback loops than the basic epidemic model, and many of these variables and loops are there to make the model more operational. These models may also have more variables and loops than are really necessary to create the underlying reference mode of behavior, but which are included either because we are not initially sure which structures create the observed behavior (our hypothesis contains multiple possibilities), or because we want the model to include more features of the real system (for example, to increase management acceptance). In our example above, the inclusion of recovered population, loss of immunity, and population growth would make the model look more like the real system, but would not affect the primary ***epidemic*** dynamics shown in Figure 1.

An operational and complete model is essential for communicating with decision-makers in the system, but as models become more complex it becomes more difficult to precisely relate feedback loop structure to behavior. Nevertheless, system dynamicists do try to relate feedback structure and behavior in order to understand why the dynamics occur and to develop and explain effective policy changes. They use a number of tools such as parameter sensitivity, loop knock-out analysis, Eigenvalues and others to determine which loops drive the behavior at different times, and thereby improve intuition about the connection between structure and behavior. Often in explaining the dynamics produced by the simulation – the relation between structure and behavior – they will use simplified stock/flow and/or causal loop diagrams, rather than the complete structure to focus on the essential feedback relationships. So, experienced system dynamicists will usually start with simple causal diagrams focusing on loops to develop a dynamic hypothesis, then develop complex operational simulation models with many feedback loops, use various techniques to understand what causes the dynamics, and finally simplify the model (diagram) again to explain those dynamics to policy makers.

If the aim of system dynamics pedagogy is to help beginners understand that system dynamics focuses on how feedback structure creates behavior, then beginning models should focus as clearly as possible on structure and behavior, so that beginners can build their own intuition

about how feedback systems work. Beginners often have a tendency to hook variables together without giving much thought to the function of each individual link or to the overall feedback structure of the model. Beginners also tend to include variables and loops because they are part of "the system," not because they are critical to the reference mode of behavior. In many cases these tendencies occur because students have not first developed a dynamic hypothesis that is an initial theory of the structure that is believed to create the observed behavior. Certainly models should be operational – it is not good system dynamics practice to lump many variables into one equation. However, at the same time, the primary goal for beginners should be to gain intuition on how structure creates behavior through an emphasis on feedback loops. For beginners, we should start with simple models focusing on the loops that create the reference mode of behavior, and then add complexity as knowledge and understanding are gained.

Therefore, in answer to our original question about which epidemic model we should use with beginning audiences and K-12 students, the two-loop version shown in Figure 2 is the best option because it clearly shows the two feedback loops that drive the behavior observed in the reference behavior mode (Figure 1). The contact rate can be included in either loop depending on the model's purpose. In personal correspondence, George Richardson points out that his introductory course model has contacts in the Depletion loop to consider the effects of vaccines. We include it in the reinforcing Contagion loop because it explains to students why they should stay home when they have the flu. In either case, the flow equation is the same, and, in reality, neither healthy nor infected people solely initiate contacts – people just mix together in the course of other activities.

The one-stock version in Figure 8 also clearly focuses on the feedback structure and is equally correct, although showing stocks for both the infected and healthy populations makes the model easier to understand visually. Both the two-stock version in Figure 1 and the one-stock version in Figure 8 teach beginners that the contagion and depletion feedback loops are necessary and sufficient to create the S-shaped pattern of behavior. However, the two-stock conserved flow better expresses students' experience with the classroom game as they move from being healthy to being infected. Also, the conserved flow prepares students to extend the model to include recovered people and other subsets of the population.

The equally correct four-loop model in Figure 6 is more operational than the two-loop version because it clarifies the calculation of the total population, but, for beginners, the added non-dynamic loops obscure the focus on the dynamic contagion and depletion loops. More advanced modelers would understand the distinction and also recognize that the two added loops through the constant population exactly balance each other producing no net change in the dynamics. However, this distinction is unnecessarily confusing for beginners and better introduced later. (Following our discussions on these models, Jim Hines reports in personal correspondence that he has used the two-loop version in introductory workshops while reserving the four-loop structure for more complex models.)

Finally, the three-loop version in Figures 4 and 5 is not correct and should not be used for two reasons. First, the third non-dynamic loop obscures the focus on the two essential dynamic loops, while it also is not balanced out by the fourth operational loop in the four-loop version. Second, the three-loop version confuses the basic purpose of system dynamics:

- Understanding how structure creates behavior in a system
- Building an intuition about how feedback loops behave
- Using that intuition and a deep understanding of that particular system to design solutions to the problem
- Transferring the understanding to other similar systems.

Unfortunately, many beginning lessons, particularly in K-12 curriculum, have drifted from this central purpose. We suggest that K-12 curriculum, and the accompanying models, can be significantly improved by following standard system dynamics modeling practice. At a macro level, standard practice dictates: (1) first formulating a dynamic hypothesis to explain the observed behavior as the basis for initial model development (rather than building models by "hooking stocks and flows together," or because a variable is in the system); and (2) once a model is constructed, conducting many simulation experiments to fully understand how the structure and parameters affect behavior. At a micro level, standard practice suggests avoiding multiple algebraic expressions within variables. We hope that this closer look at the widely used three-loop model and its better alternatives will inform improved pedagogy and understanding for beginning students at all levels.

Acknowledgements

We would like to thank Jim Hines, John Lyneis, Mohamed Mojtahedzadeh, and George Richardson for their participation in conversations and email exchanges about the ideas in this paper.

References

Glass-Husain W. 1991. Teaching System Dynamics: Looking at Epidemics, MIT SDEP Road Maps, D-4243-3, http://sysdyn.clexchange.org/sdep/Roadmaps/RM5/D-4243-3.pdf

Hines JH, Lyneis JM. 2005. Lectures Slides for SD550, System Dynamics Foundations: Managing Complexity, WPI, Session 7: 56-64.

Mojtahedzadeh M, Andersen D, Richardson G. 2004. Using *Digest* to implement the pathway participation method for detecting influential system structure. System Dynamics Review **20**(1): 1-20.

Quaden R, Ticotsky A, Lyneis D. 2004. *The Shape of Change.* Creative Learning Exchange: Acton, MA: 51-64.

Quaden R, Ticotsky A, Lyneis D. 2007. *The Shape of Change: Stocks and Flows.* Creative Learning Exchange: Acton, MA.

Sterman JD. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World.* McGraw-Hill: New York: 301.