Introduction to the Loop Impact Method

of Feedback Dominance Analysis

Session 1

Using Loop Impact to Explore System Behaviour

Notes

John Hayward

School of Computing & Mathematics

University of South Wales

john.hayward@southwales.ac.uk

sociomechanics.com

Sociomechanics

38th International Conference of the System Dynamics Society 2020

Abstract

A central premise of system dynamics is that dynamical behaviour can be explained by model structure, especially its feedback loops. For example, the familiar S-shaped growth of the limits-to-growth archetype is explained using shifting loop dominance. Although such an explanation appears clear, what is less clear is how a loop is quantified, how dominance is defined, and what aspect of dynamical behaviour is being explained. These issues become more pronounced as the number of variables and feedback loops increases.

This workshop will introduce participants to the Loop Impact Method of feedback dominance analysis. In this first session participants will explore a number of models using the method's definitions of behaviour, structure and dominance, and compare them with their own understanding of the model's behaviour and structure. Participants will be encouraged to work in groups and share ideas. Models will be accessed using the isee exchange web platform.

In the second session participants will learn how to implement the Loop Impact method by adapting standard models in Stella Architect.

Structure & Behaviour

First, we will examine what is meant by the two words structure and behaviour in the context of a system dynamics model. Ask yourself:

1. **What behaviour do we seek to explain?**
2. **What structures do we use to explain it?**
3. **How do we measure the effect of a structure?**
4. **How do we decide which structure is dominant?**

Let us specifically examine the limits-to-growth model. An interactive version is available on isee Exchange:

https://exchange.iseesystems.com/public/john-hayward/limits-to-growth-archetype/index.html#page1

Structure

A fixed amount of land is designated for a new trading estate in order to encourage business development in a town. Initially, the trading estate grows rapidly as more businesses attract more business developments - the urban attractiveness hypothesis R0. However, as growth continues, land availability falls, and business construction is reduced, B1.

The SD diagram is:

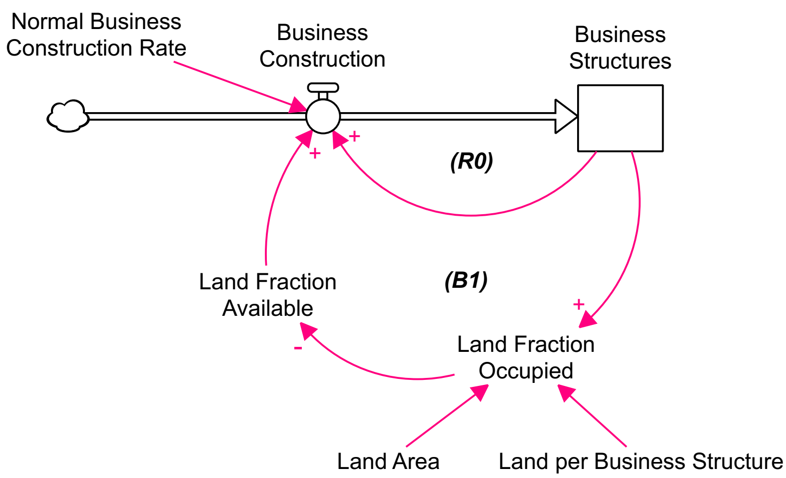

Figure 1: Limits-to-Growth Model

Structure:

- 1 stock
- 1 flow
- 2 feedback loops: 1 reinforcing, 1 balancing
- 3 parameters
- 1 initial condition
- Equations (hidden)

We will specifically use the feedback loops to explain the behaviour of the stock. Thus at least one stock and flow are involved.

So, by model structure we mean feedback loops, but note that the stock and flow structure is also involved.

Behaviour

A simulation run of the model for specific parameter values gives the S-shaped time graph for the Stock *Business Structures*:

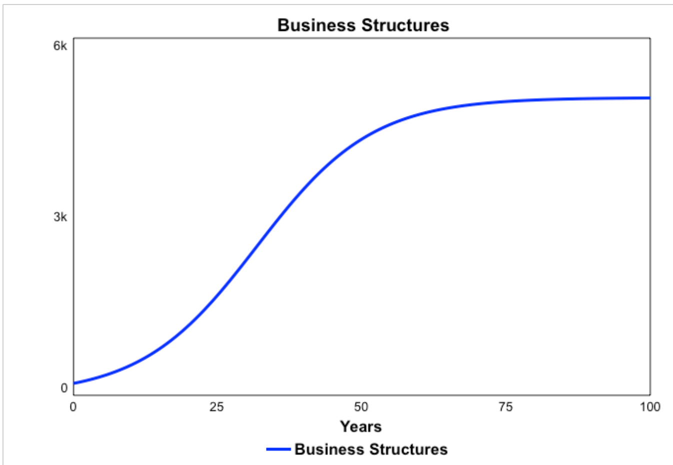

Figure 2: S-Shaped Behaviour of Limits-To-Growth Model

Notable features of the behaviour:

- *Business Structures* is either increasing or stationary.
- *Business Structures* starts by curving up, speeding up, accelerating; then curves the other way, slowing down, decelerating.
- *Business Structures* reaches a fixed limit. Other simulations show the limit is independent of the initial value.

The behaviour we seek to explain is the <u>curvature</u> of the time graph of the stock, within the context of whether it is increasing or decreasing.

In the above we would really like to know why it speeds up and slows down.

If we want to understand increase or decrease, then we would need to examine the balance of the flows.

If we wanted to determine the limit, then we would need to look at the equations.

Feedback loops are primarily responsible for curvature in stock behaviour. Acceleration (speeding up) is a reinforcing effect (R) (+), and deceleration (slowing down) is a balancing effect (B) (-). To measure the effect of a structure on behaviour we need to be able to identify loops with curvature in the time graph. That is, identify which loop structure causes a particular curved behaviour.

In models with one stock this identification is usually straightforward. Thus, for the limits-to-growth model we would hope for an explanation as in the diagram:

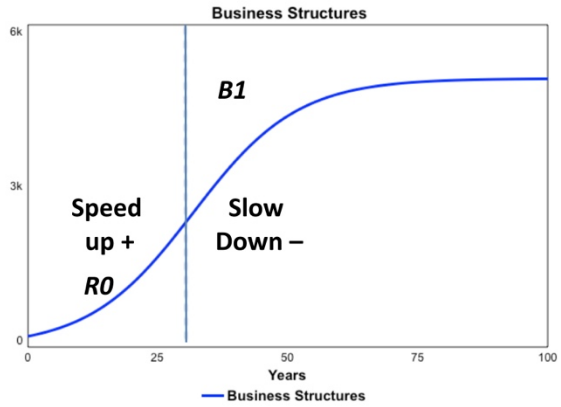

Figure 3: Mapping Loop Structure to behaviour in Limits-to-Growth Model

The reinforcing loop *R0* dominates the acceleration, whereas the balancing loop *B1* dominates the slowing down. Thus, a **cause**, i.e. the feedback loop of 2 types R and B, is identified with the **behaviour**, i.e. curvature, of 2 types: acceleration and deceleration.

Summary So Far

1. **What behaviour do we seek to explain?**
   Stocks – Curvature (speed up, acceleration +; slow down, deceleration -)

2. **What structures do we use to explain it?**
   Feedback Loops (Reinforcing, Balancing)

Identifying Structure with Behaviour

Our next question is:

3. **How do we measure the effect of a structure?**

We will answer this in two parts. Firstly:

How can we identify an individual loop with a given curvature?

This was straightforward in the limits-to-growth model as it was first order, one stock. Look at different graph that has more changes of curvature than figure 3:

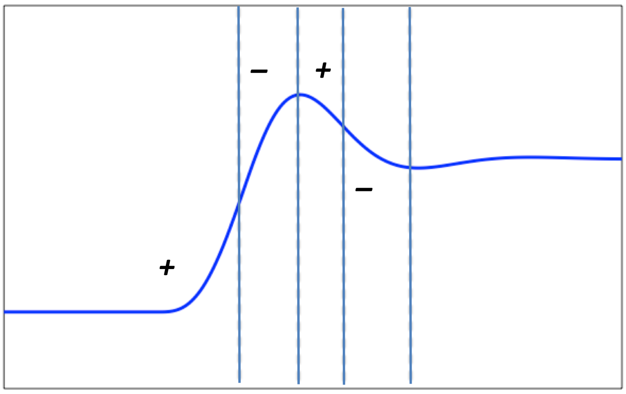

Figure 4: Understanding more complex behaviour

The above starts by increasing, speeding up, i.e. acceleration (+), then slows down (–), though still increasing. After the maximum it is decreasing, accelerating downwards (+). Then slows down (–) still decreasing.

<u>The labels only refer to the effects of the loops on the stock, not loops themselves</u>. Although the + is a reinforcing *effect*, and the – a balancing *effect*, this does not mean loops of those names cause the behaviour. The graph in figure 4 oscillates. If this is produced endogenously then the model must have at least 2 stocks, and there must be at least one second-order feedback loop, that is a loop with two stocks[1]. Such loops do not behave like the first-order loops as in, for example, the limits-to-growth. A first-order balancing loop always decelerates a stock (assuming no other loops or exogenous effects). However, a second-order balancing loop will decelerate one stock while accelerating the other. A simple example will show this.

Second-order Effects

Two stocks A and B are linked by a single balancing loop *B1*, figure 5. As a balancing loop it must contain an odd number of negative polarities. In this case the negative polarity is B's flow. Thus, at any given time, one stock will be accelerating, and the other decelerating. A simulation should illustrate this.

[1] The result that first-order, one stock, endogenous systems do not oscillate assumes time is modelled continuously. If time is discrete first-order systems can oscillate, though the oscillations are themselves discrete. Feedback in discrete systems is beyond the scope of this workshop and method.

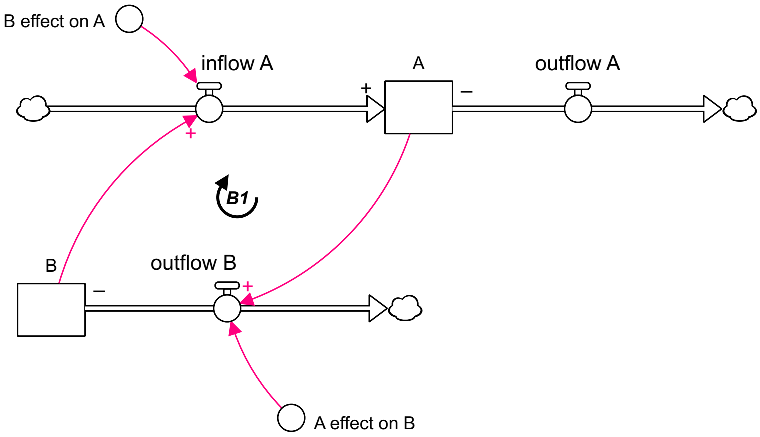

Figure 5: Single second-order balancing loop

To understand the simulation, note that both stocks are non-negative, all flows are uniflows, and A has an outflow, not in a loop, to ensure that it is able to decrease as well as increase. Initially A is set to be zero, and B is set to 10. B can only decrease, and A will start by increasing, figure 6.

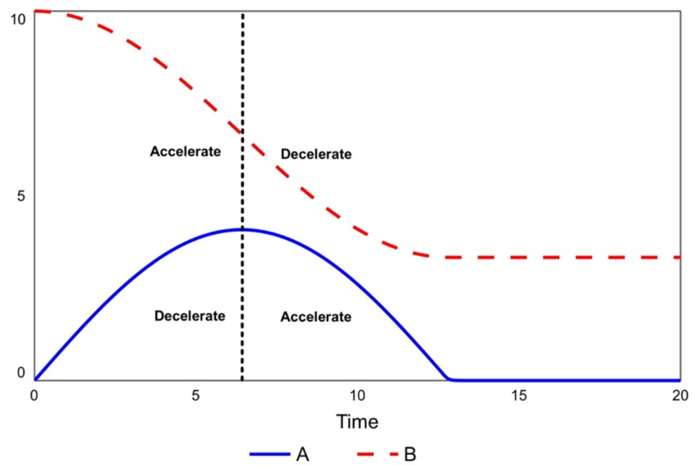

Figure 6: Effect of single order 2 balancing loop on its stocks

A starts by increasing in value, slowing down its rate of change, while B is accelerating and decreasing. When A reaches its maximum, it starts accelerating, but decreasing in value. B although still decreasing is now slowing down. In the initial phase *B1* has a reinforcing effect on B and a balancing effect on A In the second phase, after A's maximum, the effects are reverse, reinforcing on A and balancing on B. (The simulation ends due to the non-negative constraint on A.)

Given there is only a single loop it would be unwise to label the phases with *R* and *B*. Instead the labels + and – can be used indicating whether the loop is accelerating the stock (+) or decelerating it (–), figure 7. Their product always equals the polarity of the loop. Thus, because this behaviour is caused by a balancing loop, the polarity on A is always opposite to that on B. The meaning of these polarities will become clearer in the next section.

<u>Note that the one second-order loop affects two stocks</u>. Thus, thinking of the question as to how we measure the effect of a structure, i.e. a second-order loop in this case, then we need <u>two measurements</u> one each for the two effects on the two stocks.

A loop with *n* stocks will need *n* numbers to measure its effects on the stocks.

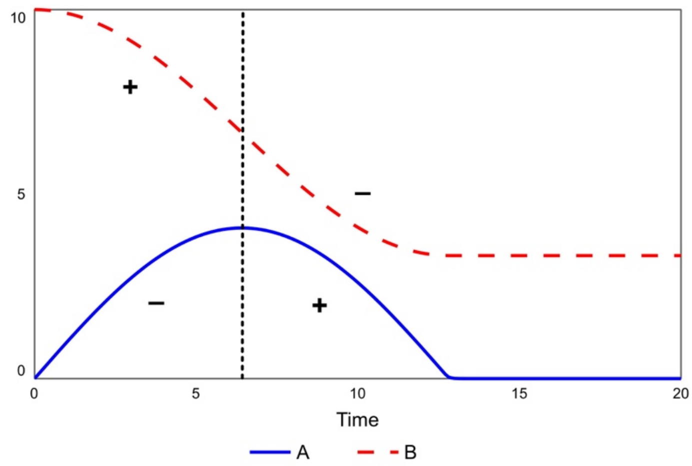

Figure 7: Effect of single order 2 balancing loop on its stocks with polarity labels

Oscillations

If the stocks are allowed to go negative and the flows in the loop *B1* are allowed to be biflows, then the behaviour of the system is one of indefinite oscillations, figure 8, with continual change between acceleration and deceleration polarities. A oscillates around zero, but B oscillates around a non-zero value due to the fixed outflow on A. Loop *B1* dominates throughout, despite the changes of behaviour.

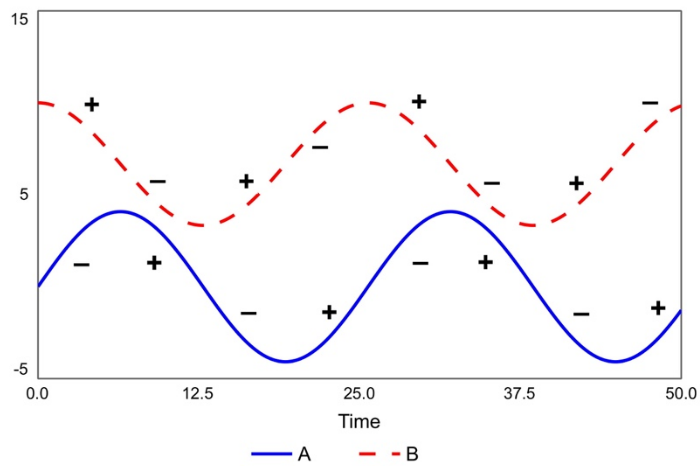

Figure 8: Oscillations caused by single second-order balancing loop

Back to the sub question:
How can we identify an individual loop with a given curvature?

It depends whether the loop is first order or higher order. First-order loops identify with a single type of curvature, but higher order loops will need to be described by their effect on each stock in their loop.

Impact

Back to question 3

3. **How do we measure the effect of a structure?**

We will now be more specific:

How can we measure the effects of the loop on each of its stocks?

Given that we have to be careful how we relate loops of two or more stocks to graph curvature, we now need a measure of acceleration in such a way that + means accelerate, and (–) means decelerate, as in figure 9.

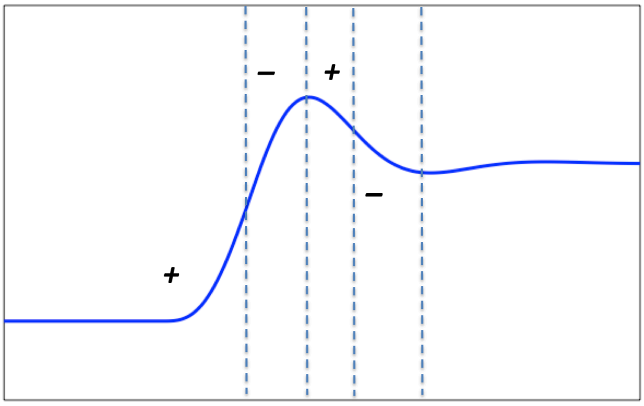

Figure 9: + means accelerate, – means decelerate

First look at the polarity of the rate of change, figure 10 (pointed to by arrow):

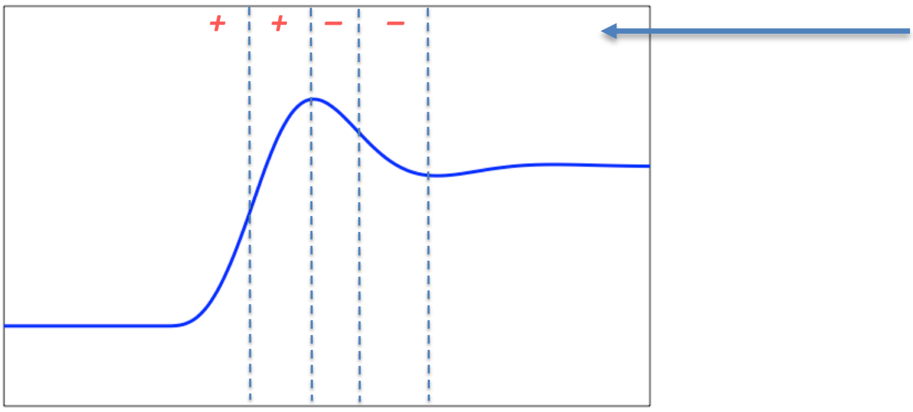

Figure 10: Polarity of rate of change, i.e. slope, first derivative, velocity.

The first two phases are positive, the curve is going up; the second two phases are negative, the curve is going down. The rate of change is the first derivative in calculus, measuring the slope of the curve.

Now consider the polarity of the acceleration, figure 11 (pointed to by arrow):

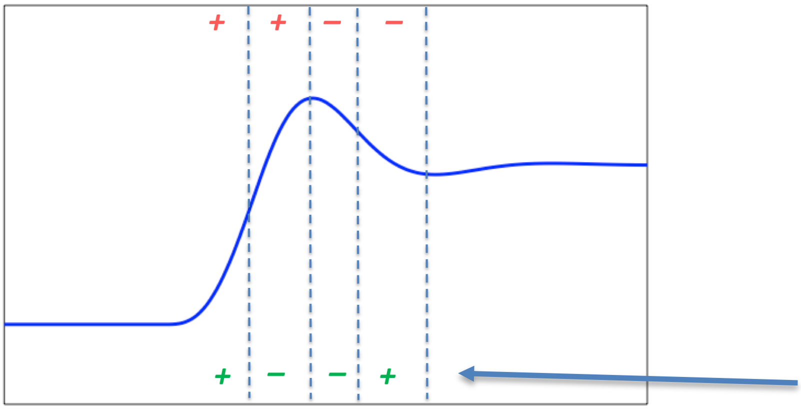

Figure 11: Polarity of acceleration, second derivative, curvature.

Acceleration is the same as the second derivative in calculus and is a measure of curvature. The first phase is positive. In the second and third phases acceleration is negative. This is best understood by analogy. Imagine there is a force pulling a body back. Phases 2 and 3 represent that pull-back, firstly slowing it down to a halt, phase 2, then bringing it back, phase 3. Because this is opposing motion, it is negative.

If you have familiarity with some calculus, then the second derivative is negative around a maximum, phases 2 and 3; and it is positive around a minimum, phase 4 and beyond. That would also give the signs.

These are not the signs we require, compare figure 9. The first two phases are correct, but the third and fourth are wrong.

To get the correct polarity for reinforcing and balancing effects we divide the acceleration by the rate of change. This is a measure called **Impact**

$$Impact = \frac{Acceleration}{Rate\ of\ change}$$

Performing the division gives the impact polarities, figure12, which now matches our intended understanding, figure 9. Impact is a ratio measure.

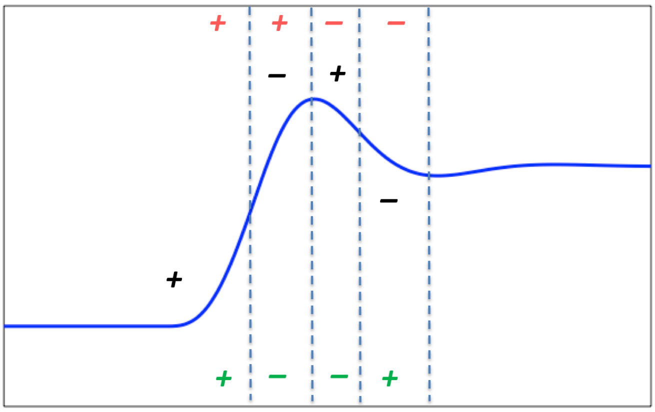

Figure 12: Impact (middle signs) is acceleration (bottom)/rate of change (top).

Thus, for the model with a single balancing loop, figure 5, its behaviour, figures 7 and 8, is explained by the impact of the loop on each of the stocks changing over time. This is typical behaviour for loops with two or more stocks. Sometimes they have positive impact, sometimes they have negative impact. But the product of the impacts in the loop is always **negative** in a **balancing loop**, as in figure 5. Likewise, the product of the impacts in a *reinforcing* loop is always *positive*.

Even a first-order model, such as the limits-to-growth one, figure 1 has its behaviour explained by impact, figure 3. A first-order reinforcing loop always has positive impact on its only stock, the first phase. A first-order balancing loop always has negative impact on the stock, second phase.

Of course, impact is more than a polarity but an actual value. Thus, for the limits-to-growth model, the impacts of the two loops can be isolated from each other, computed and plotted on a graph, figure 13[2].

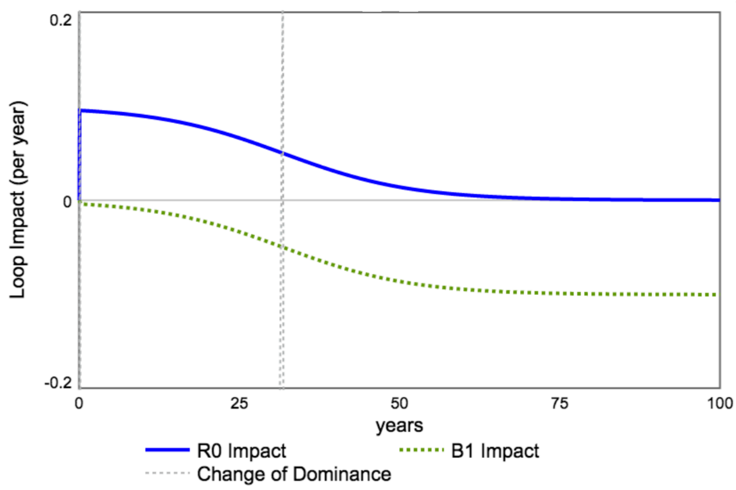

Figure 13: Numerical values of loop impacts of R0 and B1, with position of change of dominance

Note that the impact of *R0* is positive and that of *B1* is negative. **Note also that both loops contribute to the behaviour throughout.** In the first phase the impact of *R0* is bigger than that of *B1*. Thus *R0* is dominant. In the second phase *B1* is numerically the larger in impact, thus it is dominant.

That the impacts of the two loops change over time is due to the non-linear nature of the model (the stock is multiplied by itself in one of the formulae!). If a first-order loop is linear then its impact is always constant. There will be examples in the practical exercises.

[2] The method for isolating and computing the impact of a given loop on a given stock is examined in the second workshop.

For the second-order model, figure 5, there is only one loop, measured by two impacts, one on each stock, figure 14:

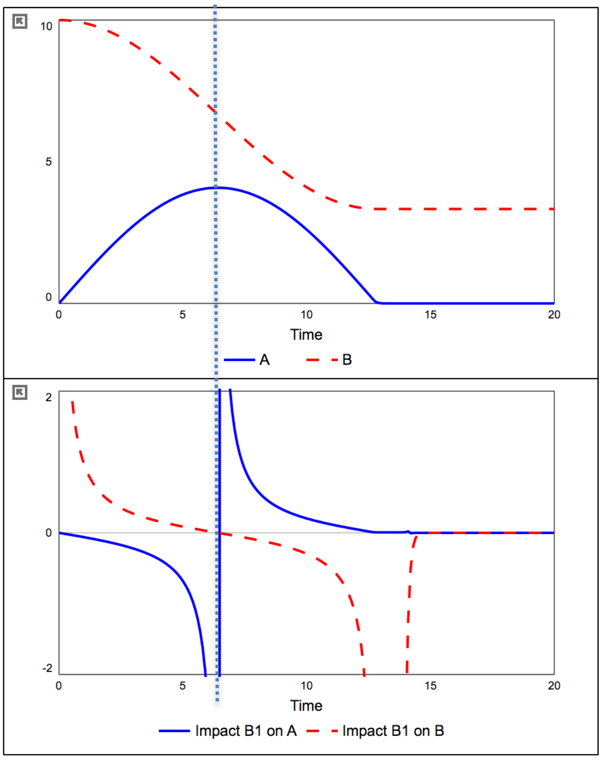

Figure 14: Behaviour and impacts of loop B1 of second-order loop model, figure 5.

In the first phase, where A is growing but slowing down, the impact of the loop on A is seen to be negative (balancing effect) and increasing in numerical value. The closer A gets to its maximum, the larger (numerically) the impact of the loop on it becomes. This is because impact is a ratio, acceleration compared with rate of change that is now much smaller. B is accelerating and has positive impact (reinforcing effect).

At A's maximum the polarities of the two impacts flip, impact on A becomes positive and impact on B becomes negative. The impact on A has changed from a balancing effect to a reinforcing effect.

It is important to grasp that a loop of second order and above can have both reinforcing and balancing effects on the same stock at different times, even though the loop is of one type.

Thus, the second-order balancing loop *B1* has a balancing effect on stock A (negative impact) followed by a reinforcing effect (positive impact). However, the corresponding effect on stock B is reinforcing followed by balancing. At any time, the product is negative – thus preserving the balancing nature of the loop as a whole.

More generally **the product of the loop impacts on all the stocks in a loop equals the loop gain**[3]

If the two impacts are measured at any point, their product is the loop gain. In figure 14 the parameter *B effect on A* is 0.3, and *B effect on A* is 0.2. Thus, the loop gain is 0.3 X -0.2 = -0.06. (the second number is minus, -0.2 because there is an outflow on B). Read the impacts at, for example, time = 5

[3] Proved in the appendix of Hayward & Boswell (2014).

Impact loop *B1* on A = -0.679
Impact loop *B1* on B = 0.0883

-0.679 X 0.0883 = -0.06 = loop gain

Although the loop gain is constant throughout, at any given point in time it is being “shared out” between the two impacts the loop has on the two stocks. At A’s maximum the loop *B1* has zero impact on stock B (inflexion point), hence its infinite impact on A at that point.

Summary

3. **How do we measure the effect of a structure?**
   By loop impact – a ratio measure of <u>acceleration to rate of change</u> for a particular loop, that is, a particular pathway between two adjacent stocks in the loop.

Dominance

Our final question concerns comparing the impacts of different loops on a stock in order to decide which one best explains behaviour. The final question:

4. **How do we decide which structure is dominant?**

The easiest way to explain this is by examples.

Consider a system where a stock is subject to three reinforcing loops and one balancing loops. They could be loops of this type, but as we have seen above they could be higher order loops which at that moment have a reinforcing/accelerating effect or a balancing/decelerating effect. For clarity we will just refer to them as loops: *R1*, *R2*, *R3* and *B1*. We will assume *R1*, *R2* and *R3* have positive effect and *B1* negative.

Figure 15 shows the numerical values of the four loops, conveniently chosen to add up to 100. The sum of the three reinforcing loops is less than the one balancing loop *B1*. Thus, the overall behaviour will be balancing (slowing down) and *B1* is the dominant loop[4]. By "dominant" we mean it is sufficient to explain the behaviour[5]

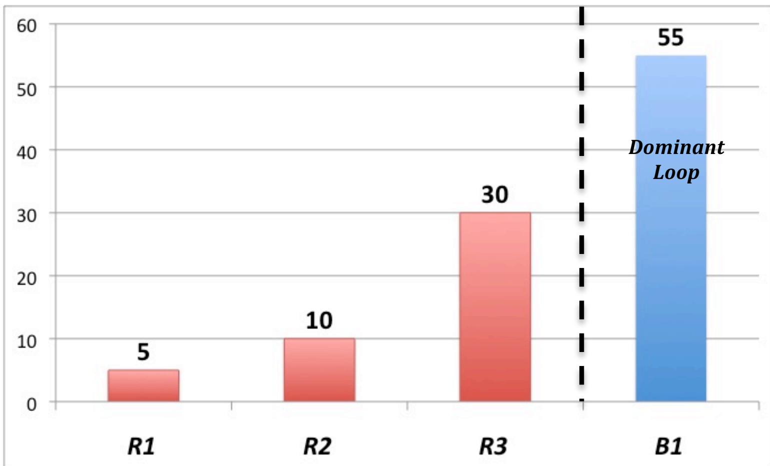

Figure 15: B1 is the dominant loop/effect as its impact exceeds the sum of the other three loop

Next, we will change the impacts so that *R3* has 55 and is thus larger than all the other loops put together, figure 16. With a reinforcing loop having 55% of the impact then the behaviour will be accelerating, and the dominant loop is *R3* as it is sufficient to explain the behaviour.

In fact, *R3* is dominant as it is larger than the single balancing loop. That fact alone is sufficient to give accelerating behaviour. It is also larger that *R1* and *R2*, thus they can be discarded for a decision on dominance.

[4] The actual impact of *B1* will be negative. But for the purposes of comparing loops, the absolute value is sufficient to make the decision on dominance.

[5] Of course, the loops *R1*, *R2* and especially *R3* do affect the behaviour, but not enough to counteract the decelerating behaviour and make it accelerate. It is the general shape of the curve that is explained by the dominant loop. The exact amount of curvature would need all the loops to explain.

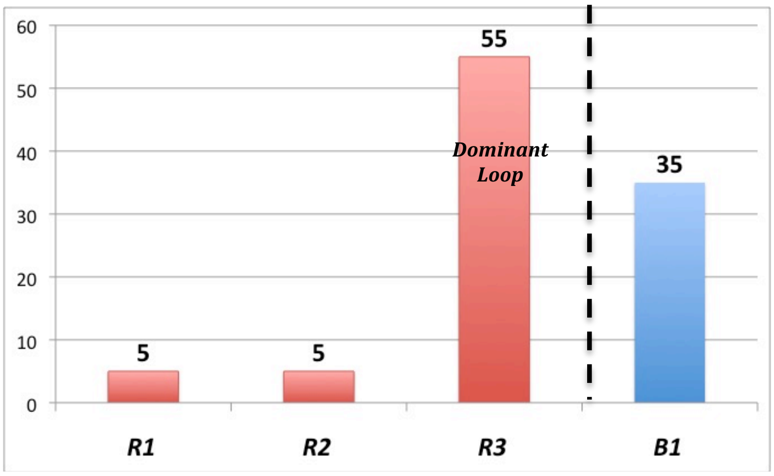

Figure 16: R3 is the dominant loop as it is larger than B1, and larger than R1 and R2

The first step in determining dominance is to decide which is the larger: the sum of all the reinforcing loops, or the sum of all the balancing loops. The winner of this contest determines behaviour. After that, it is a question of selecting out of the determining set, which loop, or loops, can be regarded as dominant.

To illustrate the difficulties that can be encountered in deciding dominance, consider figure 17. First, decide the behaviour: $R1+R2+R3 = 60$ which is greater than $B1 = 40$, thus the behaviour is accelerating, requiring a reinforcing explanation. Thus, the largest loop, in this case *B1*, is not the dominant loop – it is the wrong type as it does not match the behaviour. However, *R3* is only 35% of the impacts, thus is not sufficient to explain behaviour either.

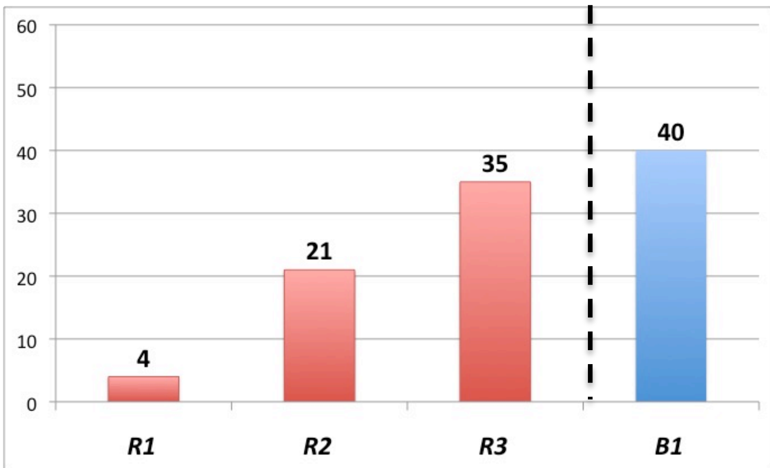

Figure 17: The case where the largest loop does not explain the behaviour

The above situation, where there is no overall winner, is rather like a multi-party election for a representative of a constituency. Using a British example, the reinforcing loops are replaced by three parties of the political left (Labour, Liberal Democrat, and Green), and one party of the political right (Conservative). The party with the largest percentage of the votes is the Conservative. However, a greater percentage of the votes were cast for left wing parties. The electorate is clearly more left than right. Who should win? It is the same problem as figure 17 which loop is dominant?

Mathematically, there is no clear way to decide a winner in an election with more than two parties. In a UK parliamentary constituency, they use the "first past the post system". That would give victory to the Conservatives, despite more people voting for the left. In some local council seats, and in some other countries' parliaments, they may use any of a number of transferable vote schemes where people make a second or third choice, or maybe put a cross on all the candidates they are willing to see elected. However useful that may be in elections, it will not help in deciding the dominant loop problem!

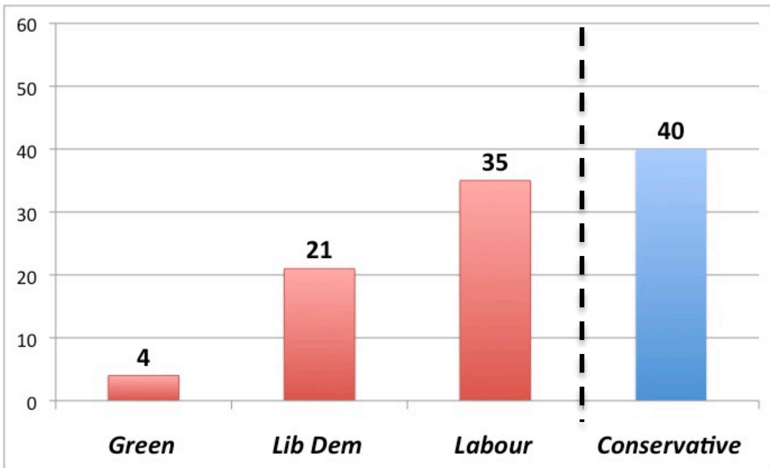

Figure 18: Percentage of Votes cast in a British Parliamentary Constituency. The left have more votes than the right, but the one right wing party is the largest. Who should win?[6]

One solution to the "no overall dominant loop problem" of figure 17 is to label *R3* the largest contributory loop, figure 19. Although *R3* is not enough to explain behaviour, it is the largest contributor to the behaviour; information that might be helpful in altering the future behaviour or dealing with a problem that needs solving.

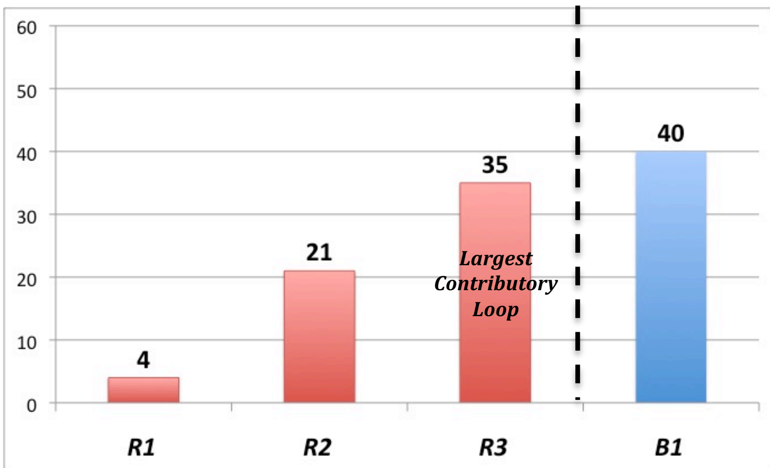

Figure 19: Largest contributory loop

An alternative way, and the one used by Hayward & Boswell (2014), is to give the dominant loop set, figure 20. This is the smallest set of loops, consistent with the behaviour, that is sufficient to exceed the sum of all the loops of the opposite type. In this case the sum of *R2* and *R3* is 56, which is larger than *B1* at 40. Thus, it can be said it takes both *R2* and *R3* to explain the behaviour. Thus, in figure 20, *R2* and *R3* are the dominant loop set.

---

[6] In the UK red is the colour of one of the left-wing parties, Labour; and blue the colour of the right of centre, Conservatives. In the USA the identification of colours with political wings is reversed!

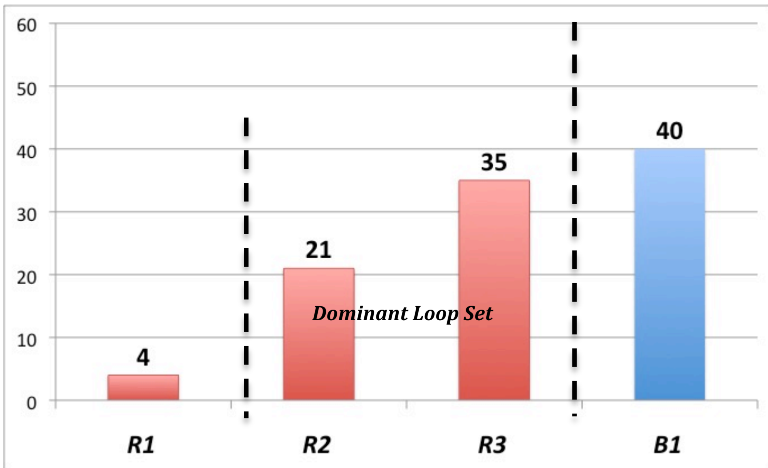

Figure 20: Behaviour explain by the dominant loop set, the minimum number of loops consistent with behaviour to achieve majority share

In figure 20 it was straightforward to work out the dominant loop set. Sometimes there are a number of choices. Consider the case in figure 21. In this case *R2*+*R3*=47 easily exceeding *B3* at 35. However, *R1*+*R3* = 43 is also is larger than *B1*, as is *R1*+*R2*=40. Which set to choose? Hayward and Boswell (2014) choose the one with the largest two loops, in this case *R2* and *R3*. This ensures the largest contributory loop is always included in the dominant loop set. The set is sufficient, but not necessary, as other choices are possible (Sato 2016).

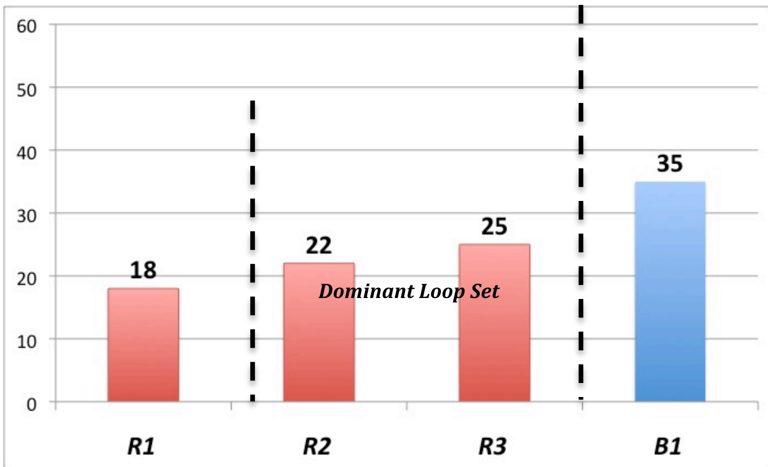

Figure 21: Where more than one dominant loop set is possible, the one with the largest loops is chosen.

Thus, when no one loop dominates, a dominant loop set is chosen. It is the set with the smallest number of loops needed to exceed the sum of the opposing loops, and uses the largest possible loops of that type.

Summary

1. **What behaviour do we seek to explain?**
   Stocks – Curvature (speed up, acceleration +; slow down, deceleration -)

2. **What structures do we use to explain it?**
   Feedback Loops (Reinforcing, Balancing)

3. **How do we measure the effect of a structure?**
   By loop impact – ratio measure of acceleration to rate of change for a particular loop.

4. **How do we decide which structure is dominant?**
   Dominant Loop Set, (alternatively: largest contributory loop)

Exercises

- Best done in groups
- Use Worksheets
- Models at isse Exchange **https://exchange.iseesystems.com/**
  - Search: loop impact
  - Select: Limits to Growth Archetype
  - Play
- Repeat for exercises

Models are written in Stella Architect by isee Systems

Models for Exercises

- Limits-to-Growth
- Gone with the Wind
- Inventory-Workforce
- Overshoot and Collapse
- Spread of Disease
- Urban Business Expansion

References

The method is described in the 2014 paper and extended in the 2018 paper.

Hayward J, Boswell GP. 2014. Model behaviour and the concept of loop impact: a practical method. System Dynamics Review 30(1): 29–57.

Hayward J, Roach PA. 2017. Newton's laws as an interpretive framework in System Dynamics. System Dynamics Review 33(3-4), 183-218.

Hayward J, Roach PA. 2019. The concept of force in population dynamics. Physica A: Statistical Mechanics and its Applications, 531, 121736,

Jeremy Sato compared Loop Impact with other methods and gave valuable insights on the method

Sato JB. 2016. State space analysis of dominant structures in dynamic social systems. PhD thesis. Washington University, St Louis, MO.

The Loop Impact method is built on previous work, particularly the Pathway Participation Metric of Mohammed Mojtahedzadeh and the Loop Deactivation method of David Ford.

Duggan J, Oliva R. 2013. Methods for identifying structural dominance. System Dynamics Review, special virtual issue. Available: http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-1727/homepage/VirtualIssuesPage.Html#methods

Ford DN. 1999. A behavioural approach to feedback loop dominance analysis. System Dynamics Review 15(1): 3–36.

Forrester NB. 1982. A dynamic synthesis of basic macroeconomic theory: Implications for stabilization policy analysis, PhD thesis, Sloan School of Management, MIT: Cambridge, MA.

Gonçalves P. 2009. Behavior modes, pathways and overall trajectories: eigenvector and eigenvalue analysis of dynamic systems. System Dynamics Review 25(1): 35–62.

Güneralp B. 2006. Towards coherent loop dominance analysis: progress in eigenvalue elasticity analysis. System Dynamics Review 22(3): 263–289.

Huang J, Howley E, Duggan J. 2009. The Ford method: a sensitivity analysis approach. In Proceedings of the 27th International Conference of the System Dynamics Society. System Dynamics Society: Albuquerque, NM.

Kampmann CE. 2012. Feedback loop gains and system behaviour (1996). System Dynamics Review 28(4): 370–395.

Kampmann CE, Oliva R. 2008. Structural dominance analysis and theory building in system dynamics. Systems Research and Behavioral Science 25(4): 505–519.

Mojtahedzadeh M. 2008. Do parallel lines meet? How can pathway participation metrics and eigenvalue analysis produce similar results? System Dynamics Review 24(4): 451–478.

Mojtahedzadeh M, Richardson GP. 1995. Confusion in the polarity of major loops. In Proceedings of the 13th International Conference of the System Dynamics Society, Tokyo, Japan. Albany, NY: System Dynamics Society.

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

Moxnes E, Davidsen PI. 2016. Intuitive understanding of steady-state and transient behaviors. System Dynamics Review 32(2): 130–155.

Richardson GP. 1986. Problems with causal loop diagrams. System Dynamics Review 2(2): 158–170.

Richardson GP. 1995. Loop polarity, loop dominance, and the concept of dominant polarity. System Dynamics Review 11(1): 67–88.

Richardson GP. 1997. Problems in causal loop diagrams revisited. System Dynamics Review 13(3): 247–252.