RT2: Systems Dynamics applied to Behavioural Operations

Harry Newton

01/04/24

1 Introduction

At present, a majority of Operational Research theory and methodology assumes that decision makers, as a whole, are rational and make choices in order to pursue an optimal outcome. In reality, humans still make up a vast majority of the decision makers in business, institutions and governments. For a long time we have known that humans are not completely objective - limits in our cognition as well as external factors such as stress and time, lead consistently to sub-optimal decisions. Additionally, within the same organisation, management styles are heterogeneous, meaning that the reasoning behind management decisions vary based upon the personality, cognitive abilities and experience of the individual in that role (Goldfarb and Yang, 2009). Our cognitive limitations with regard to large and complex systems can also result in undesirable outcomes. Studies related to human conceptions of accumulation and over-exploitation of resources have show that we do not intuitively comprehend nonlinear effects - exponential growth and decay (Moxnes, 2000).

Therefore, human reasoning has been identified as **boundedlly rational** - we reason towards an *optimal* decision, but the rationale behind our choices is bounded by our **cognitive limitations and biases** (Bendul and Knollman, 2016). We constantly experience a deluge of information, which without some kind of sorting procedure would quickly overwhelm us. So our brain applies simple rules of thumb to pick out data which appears relevant or agreeable to us. This is called the *framing effect*, and explains *confirmation biases*, where we favour information which suites and justifies an already held belief. In contrast, a rational agent has no bias toward any particular piece of information, they have an *invariance of preferences*. In stressful and busy industries, decisions need a quick turn around, limits to human cognitive processing lead decision makers to *satisfice* an objective rather than spend the time they do not have to develop a more optimal strategy. To *satisfice* means that the decision maker seeks to fulfil the minimum possible requirements to reach a goal. Neglecting these behavioural effects in our models can lead to undesirable results, or for methodology in industry, the possibility for individuals to discard(Duchessi and O'Keefe, 1992) or work around the methods entirely(Pollock, 2005).

The field of *Behavioural Operations* tries to correct for this shortcoming in classic OR theory and practice, by attempting to study human behaviour and cognition and its impact on operating systems and processes (Gino and Pisano, 2008). One idea behind it being that with a greater understanding of human behaviour, we may be able to better tailor our methodology to be more suitably applied and effective in human systems.

2 Behavioural Operations

In general Behavioural Operations Research (BehOR) is a broad field with no formeal integration, although centred around OR much of the theory has been borrowed Behavioural Economics, Cognitive Psychology, Game Theory and Systems Dynamics among others(Martin et al., 2016). There is a large component of Behavioural Operations dedicated to empirical research; its importance in this field is self evident, but we will not cover this in our report. Those who are interested can look at chapter 4 in (Ibanez and Staats, 2019). For an example of empirical Behavioural Operations research in practice, one can look here (Goldfarb and Yang, 2009). As for theory, there has been many attempts to model and understand the behaviour of individual decision makers. Primarily using methods from Behavioural Economics and Game theory, Prospect Theory is one of the more well known examples of this. Most behavioural models are structured around the *utility function*, which maps a number of choices e.g. money, products etc. to a particular utility value. The idea being that a decision maker will make a choice or a series of choices which maximise their utility. The goal then is to build this utility function in a way which reflects the behaviour of a decision maker. Other variants of this model introduce fairness (Rabin, 1993), or risk sensitivity (Prospect theory) (Kahneman and Tversky, 1979). Humans are more sensitive to losses than we are to gains, if a choice means we have a greater risk of losing something, even if we can gain more from the gamble; we are, on average, less likely to take that risk.

Systems Dynamics (SD) is another modelling framework which can, and has been applied to study decision behaviour in real world systems (Oliva and Sterman, 2001)(Morrison, 2015). The framework's effectiveness stems from its ability to directly incorporate difficult to model - emergent behaviours, which are features of large complex systems. Unfortunately, except for a few standard bearers (Sterman, 2009)(Hämäläinen et al., 2013)(Morrison and Oliva, 2019) there has not been many studies using SD to model behaviour. This is likely down to a number of reasons for this, which we will discuss later (see Section 4). In this report we intend to demonstrate the potential and power SD has in providing more detailed descriptions of human behaviour. How it can be used with informed judgment to infer the decision rules managers apply in response to different problems (see Section 3.3). We will begin by first discussing some of the important features of complex systems (Section 3). We will then introduce the methodology used in SD to formally describe systems for use in modelling (Section 3.1 & 3.2). We will then discuss how data is collected to inform model formulation (Section 3.4). Then we will briefly cover the methods used to validate these models (Section 3.5). Finally, we will discuss future research and the limitations of this approach (Section 4).

3 Modelling Dynamic Systems

Systems are inherently dynamic and nonlinear, they are constantly adjusting and adapting in response to inputs and stimuli. A business for example, must keep on generating revenue, if one supplier is cut off or goes bankrupt, the managers within the system will quickly secure a new one. If they do not, the business collapses and they lose their jobs. The recourse afforded to these managers is *path dependent*. Any previous decisions or events imposed upon the system limit or expand the choices available to the decision maker (DM). A simple example would be a company which wants to expand into a new market. The extent of the companies planned developments is dependent on prior market research, its current profits, recruit availability and the current state of the market itself, among other things. *Path dependence* also states that most decisions, once implemented are *irreversible*. Meaning, that if one wishes to backpedal on a decision, then the resources, capacity and time already invested into its realisation cannot be undone. Systems are also *irreducible*, so that the individual components of a system cannot be isolated or removed while still retaining their previous behaviour when part of a whole. If we want to study the behaviour of a single manager or *agent*, then we must do so in the context and conditions in which they normally act. The manufacturing department cannot exist without a sales department to generate revenue, they both depend on each other and the greater organisation; they are said to be *emergent* properties of the company system.

3.1 Formulating System Structure

The complex and dynamic interactions we have just mention inherent to systems mean that mathematical formulae alone cannot provide a sufficient or intuitive description of a systems behaviour. We must build another representative language on top of mathematics to help make our models more comprehensible to ourselves. We are also designing our models for the purpose of understanding the operational problems among real world systems. Meaning, that within the modeling process we are interacting and working with multidisciplinary teams of researchers and decision makers, inside and outside of these systems. Of course, not all of our colleges have technical backgrounds, so this extra language must be easily transmittable without too much prior expertise.

As such, a visual representation provides a more intuitive description than mathematics alone. The *stock-flow diagrams* and *causal loop diagrams*, seek to capture the more unintuitive, complicated parts of a system and represent them in a more digestible format. In this section using content from ste (2009a) and ste (2009b), we will provide a brief introduction to these model formulation methods, and explain how they can be used to show the concept of *feedback* and the movement of resources throughout the system.

3.1.1 Stock-Flow Diagram

At its most basic level, a system can be seen as the movement, or *flow* of resources and/or information. Our circulatory system is a prime example of this principle, the heart pumps oxygenated blood around the body to keep our organs functioning. It must also pump deoxygenated blood back through the lungs for re-oxygenation. The *flow rate* at which the

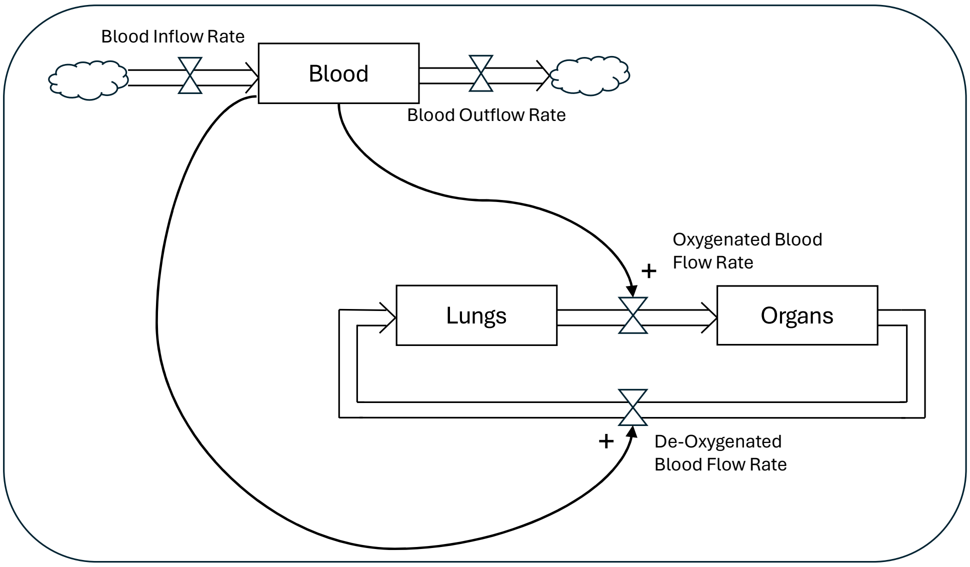

Figure 1: A stock flow diagram of the cardiovascular system. Showing the flow of oxygenated blood to and de-oxygenated blood away from the organs. Shown also is the causal relationship between blood levels and blood flow rate (blood pressure).

heart pumps our blood around the body is our blood pressure. In this visceral example, the lungs and organs are represented as *stocks*, which are the labeled rectangles in Figure 1. We have also included a more abstract *stock* for the total blood within the system.

Stocks can hold/represent any quantity or un-described sub-system. *Stocks* can also store information, or other abstract concepts such as customer satisfaction or worker pressure. In principle *stocks* represent some quantity the system uses.

The unidirectional pipe connecting the *stocks* represent the *flow*; in our circulatory system these are the arteries, veins and capillaries. The hourglass shaped symbol on these *flows* is called a *valve*- representing the *rate* which material flows, this is our blood pressure. Finally, the cloud like symbols attached to the 'blood supply' are a *source* (input) and *sink* (output). For our purposes, we do not need to include the exact mechanisms which deal with blood creation and expulsion, so we treat these flows as *exogenous*. Put in another way, these sub-systems lie outside of our model's *boundary*, so that our model is considered to be an *open-loop* system.

3.1.2 Causal Loop Diagram

The very simple model we have just described implicitly assumes a system in *equilibrium* i.e. given a constant input, the exact state of the system - its *rates* of *flow* and *stock* levels will all remain constant. It would be useful to ourselves, and our doctors, to see what occurs if some external influence disrupts our models harmony, so that it falls into dis-equilibrium.

For example, smoking and foods high in fats can fill our arteries with plaque, which

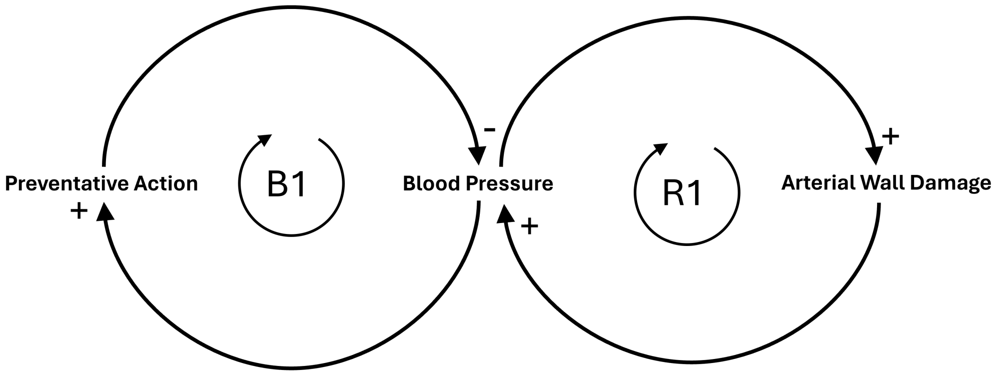

Figure 2: A *causal loop diagram* showing the relationships between blood pressure and cardiovascular damage. These form the reinforcement loop $R1$. Preventative action and blood pressure are anti correlated and form the balancing loop $B1$

forces the heart to pump at a faster *rate*; damaging the linings of our arteries. As a result, blood and plaque gathers and *accumulates* at the injury sight. Again the heart must beat faster, and still damage is incurred, until a complete arterial blockage leads to a heart attack. The process we have just outlined is an example of a *positive feedback loop*- a form of *path dependence* in which previous events in the system compound, forcing some *stock* to increase and accumulate. What we are describing in essence is *exponential growth*.

As we are not systems thinkers by nature, many inexplicable and undesirable outcomes from our decisions, turn out to be a direct result of some unaccounted for runaway feedback loop (Hämäläinen et al., 2013). It is therefore crucial that we include feedback mechanisms within our model.

So it is here that I introduce our next diagram method: The *causal loop diagram*. We start with a basic causal link between two variables and represent one of the variables' effect on the other with an arrow. The $+$ or $-$ label on the arrow tell us whether our cause has a correlating relationship ($+$)- if the cause changes, the effect will also change in the same direction. If cause increases, so will the effect and *vice versa*. Or an anti-correlation ($-$). If causes increases, effect will decrease and *vice versa*. These arrows cannot tell us the exact nature of these relationships; whether they are linear or non-linear, but this basic form can still provide us with useful insights about the nature of the system. If we now include a second, opposing arrow, between our two variables we have created a basic feedback loop. We can see our heart disease example as a *causal loop diagram* in Figure 2.

Feedback loops come in two distinct forms, these are *reinforcing loops* and *balancing loops*. The right-hand loop, labeled $R1$ of our heart disease example demonstrates a *reinforcing loops*. If our patient begins practicing measures to lower their blood pressure, we can include the additional left-hand *balancing loop*, labeled $B1$. As the patients blood pressure increases, they respond by adopting healthier habits, thus blood pressure goes down.

These loops can grow significantly in complexity; often including chains of variables and multiple interacting loops. Later we will see examples of this kind in a more detailed scenario. For now we will discuss how to represent these relationships mathematically.

3.2 Mathematical Representation

The diagrams we have just discussed cloak many unseen nonlinear relationships and interactions; these must be given an appropriate mathematical description to make computational simulation possible. In most of the literature these formulae are also given simple labels to give clarity when presenting to non technical audiences (Sterman, 2009)(Morrison, 2015). An integral function is denoted by $INTEGRAL$, for example. As we assume the reader of this report to have a decent mathematical background, we have chosen to stick with standard mathematical notation. Except for a few adaptations, most of the theory presented in this section can be found here (ste, 2009c)

3.2.1 Time Dependent Stock Flow (Accumulation)

Given a stock S with an *inflow* rate function $f_I(t)$ and *outflow* rate $f_O(t)$, then the quantity of stock at time T is described by its net flow and its initial level S_{t_0}:

$$S_T = \int_{t_0}^{T} (f_I(t) - f_O(t))dt + S_{t_0} \tag{1}$$

Where $t_0 = 0$ is our initial time. Although mathematically simple, human decision makers do not automatically apply this reasoning to *accumulation* problems within real systems (Hämäläinen et al., 2013). In the departmental store questionnaire, participants are given a time-series graph of two plots- one line is the varying inflow of customers for a store and the other is its outflow. The participants are asked a series of questions such as: 'At what point does the store receive the most customers' to harder ones such as 'At what time does the store have the largest amount of customers'. Even in the easier question, which can be answered simply by taking the difference between inflow and outflow or the derivative, participants, educated professionals among them will get the answer wrong. They instinctively associate the *stock* quantity to the flow at a given point, so that if the inflow is very high, then they assume that the point is when the most customers were present.

3.2.2 Flow Rates

Flow rates $\frac{dS}{dt}$ into and out of a *stock* are often related to the size of the stock. We can represent this relationship using the fractional rate variable $\alpha(t)$:

$$R(S,t) = \frac{dS}{dt} = \alpha(t)S(t) \tag{2}$$

If α is a constant then, we can describe the above relationship using a linear 1st order differential equation; in the language of systems dynamics this is a *first-order feedback system.* Simply put, when α is constant S will grow ($\alpha > 0$) and decay ($\alpha < 0$) exponentially with time constant $\tau = \frac{1}{\alpha}$.

When we begin modelling systems of significant size, we find that flow rate may be affected by quantities other than stock or time. For example, in a business the rate at which new employees are recruited can be influenced by a multitude of external and internal factors. Say the unemployment rate is high, then recruitment rate should increase as a result. However, high unemployment may suggest a recession, which will make the company reluctant to hire new workers as demand will probably be lower. We can continue listing factors *ad nauseam*, the fact is we need some way of representing these numerous, and probably nonlinear effects mathematically; without making our models and notation overly complicated.

Consider our stock's flow rate R, to have n independent, correlating variables $\{X_i\}_{i=1}^{n}$, so that $R = f(X_1, X_2, X_3, \ldots, X_n)$. We don't expect these variables to have a *homogeneous* relationship with R, nor do we expect them to share identical *units* with R, or each other. Thus it becomes necessary to *normalise* R and each X_i around some *reference point* we will denote with $(*)$. The reference point R^* might be Rs value at a 'business as usual' state, where the system lies in a partial *equilibrium*. Likewise, X_1^* is the value of that effect when $R = R^*$, as is the case for all the other independent variables. We can therefore represent R as a sum or product of all its dependent relationships:

$$R(X_1, X_2, X_3, \ldots, X_n) = R^* + \sum_{i=1}^{n} f(X_i/X_i^*) \tag{3}$$

$$R(X_1, X_2, X_3, \ldots, X_n) = R^* \cdot \prod_{i=1}^{n} f(X_i/X_i^*) \tag{4}$$

The choice of either *product* or *sum* can depended on computational tractability, or accuracy to the real world system. If food supply was set to 0 for a population model, then it is not realistic to have population increase regardless. In this case, the *product* form would be more appropriate.

3.2.3 Delays

In our models so far, we have assumed that when the causal effects experienced by a variable change, they do so instantaneously. Of course, in real systems this is certainly not the case, and the movement of goods and information around an organisation or supply chain are subject to *delays* and *lead times*. To properly capture the behaviour of *decision makers* (DM) in real systems, we must include these effects. The *bull-whip effect*, is a prime example of this reasoning. In *decentralised supply chains*, misinterpretations of feedback by human decision makers leads to massive oscillations in demand, which propagate along the supply chain. One contributing explanation of the *bull-whip effect*, identifies a disparity between the delays inherent in the system and the DMs mental model of how the system operates (J. Nienhaus and Schoensleben, 2006). As such, a DMs response to delays is a property which can lead to many undesirable outcomes, which is why its inclusion in our model is important.

It is common practice to apply *adaptive expectations* to describe how a DM perceives information. Imagine the DMs expected knowledge of a quantity at time t to be represented by the *information stock* variable $\hat{I}(t)$. The real information $I_{real}(t)$ is usually a discrete quantity which is constantly updated, and updates $\hat{I}(t)$ with a delay T_D. Older knowledge $I_{real}(t_{old})$ will have less weight in decision making in comparison to newer knowledge. We can therefore define a *first-order information smoothing* function or *exponential smoothing* function to describe this knowledge update process:

$$\hat{I}(t) = w(t)\hat{I}(t_{old}; t_{old} < t) + (1 - w(t))I_{real}(t) \tag{5}$$

Here the information weighting variable w is a decaying exponential, with delay constant T_D

$$w(t) = \exp\left(-t/T_D\right) \tag{6}$$

Material delays are another important feature of real decisions, if a manager has made an order to fulfil present customer demand, and has neglected this delay in their reasoning then they have made a bad decision.

The simplest form of delay, the *pipeline delay* defines the output $f_O(t)$ as a translation of the input, if we request an order of 10 units from a supplier and expect a 2 week delay, $T_D = 2(weeks)$ on delivery then we can include this in our model using the function below:

$$f_O(t) = f_I(t - T_D) \tag{7}$$

Pipeline delays works well for *first in first out* (FIFO) processes, but oftentimes, multiple deliveries may come in at different times; be mixed with other deliveries and so on. An example could be a number of orders requested from a wholesaler at subsequent different times from each other. Stock available to the wholesaler may be shipped immediately, even if it was ordered later. We can characterise this as a *first-order delay*, where the *net material in transit* is $(f_I(t) - f_O(t))$:

$$f_O(T) = \int_{t_0}^{T} ((f_I(t) - f_O(t))/T_D)dt + (f_I(t_0) - f_O(t_0)) \tag{8}$$

3.3 Putting it all together: Modelling: Anchor and Adjustment Heuristic

We have now covered the basics of Systems Dynamics models, we will now apply what we have learned to a common decision heuristic called *Anchor and Adjustment*. Real world managers make quick decisions under stress and time pressure. They may be able to use sophisticated OR techniques to help inform their decisions, but as often happens unforeseen events can prevent these *optimal* values from being achieved. Alternatively, a manager may need to quickly adjust some order, or make an off the cuff prediction for resource demand. They are not going to ‘pull a figure out of thin air‘, so to say, they will instead *anchor* their reasoning on some concrete value e.g. *average demand*, *previous orders* etc. and increase or decrease that value in response to system feedback.

We can describe this regulatory feedback loop as a *Hill Climbing Search*- a local optimisation heuristic, where a decision maker starts with an initial value and incrementally increases it until some objective function has been maximised. For a convex problem this will lead to a *global optima* - the best solution to the DMs problem. Our real world problems are not likely to be convex, so this method cannot guarantee an optimal solution. However, this fact makes *Hill Climbing Search* an appropriate model as a human decision rule. In fact (Oliva and Sterman, 2001) used *Hill Climbing Search* in their SD model investigating service quality erosion for a UK based bank. In response to work backlogs,

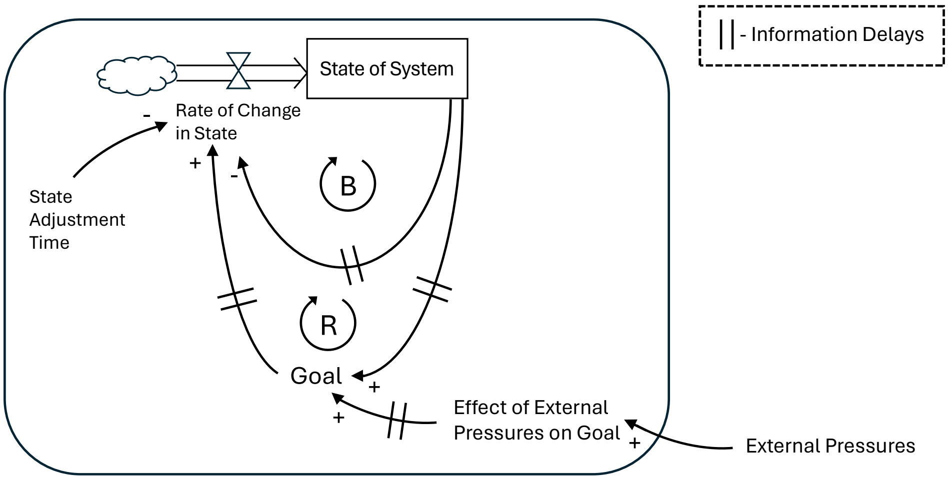

Figure 3: A systems dynamics diagram showing the *anchor and adjustment* decision rule. This model incorporates information delays, manager action and goal revision. The system adjustment balance loop is denoted B, and the goal revision reinforcement loop is denoted R. Adapted from (ste, 2009c)

bank employees would reduce the time they spent on each case, by going off the previous standard T and decreasing downward in response to *work pressure* from management.

Let us imagine our manager has *anchored* on some goal or desired system state S^*; they implement *policy* to bring the system towards this new state, by trying to reduce the error $\Delta S = S^* - S$, where S is the current state of the system. We can see this portrayed in the *first-order feedback decision loop* in Figure 3. The mathematical form of this loop is below, where $\hat{T}_G$ is the expected goal adjustment time:

$$S = \int_{t_0}^{T} (S^* - S)/\hat{T}_G dt + S_{t_0} \tag{9}$$

As the manager converges on S^*, new information about the system after their actions have taken effect may cause them to revise their goals. This may be that S^* was unrealistic, or insufficient in achieving a desirable outcome. There are many factors which can lead to S^*'s adjustment; we can use Equation 3 or 4 to formally represent these effects $\{X_i\}_{i=1}^n$. We also know from the standard formulation that S^*'s adjustment will depend on the systems current state S:

$$S^* = S \left[\prod_{i=1}^{n} f(X_i/X_i^*) \right] \tag{10}$$

To approach a closer depiction of reality, we can also include a *first-order information smooth* on the output of S and the other effect variables. In addition, when our manager actions their *policy* there will be delays in its effect. For example, production capacities may need to increase to meet this target, this will require the order of new machines and staff which will be handled by other departments. If our model boundary encompasses only this management process then we need to include delays for *exogenous* inputs. The reassignment of staff may only require a *pipe-line delay* characterisation, where as the requisition of new machinery need a first or higher order delay. For now, to keep our formulation simple we will maintain our current structure.

To simplify even further, let us now assume that the cumulative effect of the effect variables can be concentrated within the constant k. So that the relationship between the managers goal and the systems current state takes this form:

$$S^{*} = Sk \tag{11}$$

With this updated relationship we can rewrite our *first-order feedback decision loop*:

$$S = \int_{t_0}^{T} (k-1)S/\hat{T}_G dt + S_{t_0} \tag{12}$$

If we vary k we can see how the manager's *anchor and adjustment decision rule* reacts. If $k > 1$, the goal revision loop R dominates as the manager constantly re-assesses the capacity for the system to grow. If $k < 1$, then the manager is facing significant resistance to their initial goal, they then reduce their expectations and define a lower standard for their desired state of the system. The system will reach a new equilibrium once S and S^{*} converge to a single value.

3.4 Research, Field-Work & Data

We cannot create our model within a vacuum, theory is a good starting point but to build an accurate and useful model we must find out about the system itself; we must understand how the different subsystems and agents interact, for this we need good data. As we are dealing with human systems, the ideal data sources would be what Jay W Forrester calls the *mental data base* (Forrester, 1980) - The way human decision makers respond to their own observations on the system, their expectations and prior experiences with the systems behaviour and the *Decision Rules* they apply. Forrester, was one of the early pioneers of *Systems Dynamics*; he argued the importance of *qualitative* data, in addition to numerical statistics and measurements. As we cannot peer directly into peoples minds, we must build our understanding of *mental data bases* (Forrester, 1980) using *Ethnography*- The study of the culture and behaviours of a particular group. Although, *Ethnography* is usually confined to Anthropology, much of the research techniques they use i.e. written records, interviews etc. can be applied to building SD models in Behavioural Operational Research. Numerical data alone is not sufficient to describe the actual causal mechanisms behind decision behaviours; statistical and inductive inferences may be collated, but can be heavily biased if the source of these judgments is from the modeller themselves. Someone who is an outside party.

A good demonstration of field research can be found in (Morrison, 2015), in which Morrison began collecting data for his model over a period of 20 months. He started his observations just as an engine manufacturing company began implementing a new 'lean manufacturing policy' to reduce inefficiencies in the manufacturing process and help reduce work loads. His data consisted of numerous interviews with managers, employees and union officials at different points as the new policy was taking effect. He made sure to collect audio recordings of his interviews for future reference, as well as company documents and emails. During the data collection process he also frequently illustrated what he learnt with *causal loop diagrams*, to help him understand the various processes taking place within the company. Iterating between data collection, diagram creation and the creation of formal mathematical models allowed him to identify useful patterns, test them and update his conclusions. In a sense, this field work was not an isolated part of his study but part of the model creation process itself.

3.5 Model Validation

Model validation is extremely difficult, our model is usually a bespoke but imperfect replication of one particular system, so we cannot perform repeat tests and quantify its truth with a convenient statistical parameter. In ste (2009d), Sterman argues that we can never validate or verify whether a model truly represents the system under study. He suggests that such a test would be impossible, after all 'the map is not the territory' as quoted by Alfred Korzybski(Korzybski, 2005). Instead, Sterman says that we should assess the models usefulness; that is the model's ability to provide answers and explanations to the questions we are asking. He has conveniently collated many useful methods to test a models usefulness. This list is extensive, so I will only briefly cover and explain some of them in this section. For a more thorough dive into this topic, along with useful cases and explanations one can explore chapter 21 of *Business Dynamics by John D. Sterman*(ste, 2009d).

One more important point, *validation* is not the only important factor when conducting Systems Dynamics studies, one must also make their research accessible and clear enough so that others may asses and use it. To help us achieve this goal, we can take advice from the application of the **5Rs** in software engineering (Benureau and Rougier, 2018)- **Rerunnability, Repeatability, Reproducibility, Reusability and Replicability**. One notable and crucial practice is documentation. You must include the reasoning behind your models formulation, the equations you used to model relationships, your units and all the justifications which go along with these components. Describe your variables using plain English, instead of R_d, use Rate of Increasing Demand. It is also important to label your diagrams structures and loops, preferably within the diagrams themselves.

3.5.1 Structural Assessment

Some of the simplest sources of error within our models arise from behaviour which fails to conform to physical reality. If we have an inventory *stock* variable that has gone to 0, then no matter how many orders have been placed, they cannot be fulfilled. Unfortunately, we cannot manifest stock from the *aether* to make up for shortfalls. This is may seem trivially simple, but that very triviality makes such errors easy to overlook. In such cases a simple control measure should be implemented within the model to shut of a *stock's* outflow once it reaches 0.

Failure to account for irreversible processes is another simple error which can lead to amusing results. Sterman (ste, 2009d) provides a real case of a model for the leather market, which, when demand was low, would convert unsold product back into cows! Structural assessments to ensure logical and physical consistency should be performed repeatedly throughout the models construction and finalisation. They should be informed on the modeller's judgment, as well as the valuable data collected from the system itself.

Dimensional consistency is another very important factor, it is useful to check if the variables within your model actually have real life counterparts; that there is no variable with inexplicable units such as (work pressure/week^2/month^3).

The use of structural assessments are obviously important in ensuring model coherency, but as well as this, the act of performing the assessment helps the model develop a firmer grip on their understanding of the system.

3.5.2 Quantifiable Assessments

The equations within a model will all contain parameters which will need to be determined and tested from the system's real data. Depending on the nature of a variable, finding time-series data for system ranges from easy to almost impossible. Quantities which the business/organisation/system records for their own purposes may not be accessible; depending on the researchers relationship with that system, i.e. if they are conducting an independent study or have been commissioned by the organisation itself. Other 'soft' variables like work pressure, customer satisfaction are more difficult to obtain, and may require extra work on the researchers part. We can borrow from the social sciences by using questionnaires and surveys to gauge these quantities, but as there is no 'objective' measure, a combined use of modeller's judgment and statistical analysis may be required. With limited available data *Bootstrapping* methods can be used to fill gaps in the data, but care should be exercised to ensure the new data maintains the appropriate temporal behaviour.

Appropriate statistical methods should take into account the *collinearity*, *autocorrelation* and *heteroscedasticity* of the data, when estimating parameters and their *confidence intervals*. So Ordinary Lease Squares (OLS) regression methods, should only be applied if these effects are shown to have little statistical significance (ste, 2009d). Chapter 1 of *Analytical Methods for Dynamic Modelers* (rah, 2015) describes applications of parametric and non-parametric Maximum Likelihood Estimation (MLE) methods for determining parameter values and their confidence intervals for dynamic models. They also describe *Bootstrapping* methods which can be used for parameter estimation if *asymptotic assumptions* for the MLE are shown to breakdown or produce inconsistent results. They suggest that *Bootstrapping* is more computationally demanding some elements of its application are left to the modellers judgment. See (Oliva and Sterman, 2001) a real example of thorough data analysis and parameter estimation.

On that point, despite the subjective nature of that statement, a modellers judgment should not be neglected, and sometimes it is even necessary to close in on parameter values if statistical estimation is unreliable. So long as this judgment is not absurd, an unverifiable parameter bound in strong reasoning is better than no parameter at all. However, this attitude should not be applied lackadaisically to parameter estimation, and if it is done, it should be stated within the documentation along with justification.

3.5.3 Sensitivity Analysis

Generally, sensitivity analysis tests assess the systems behavioural response when parameters are adjusted within particular ranges. The purpose of this process is to test the robustness of ones' conclusions about the models behaviour; to see if they are still valid in response to adjustments in how the system operates. Additionally, it also serves to check the suitability of the confidence regions for the estimated parameters, or parameters that have been set by judgment alone. For *Sensitivity Analysis* in behavioural systems, Sterman encourages us to put less emphasis on *numerical sensitivity* (parameter response) and more on *Policy Sensitivity* and *Behaviour-Mode Sensitivity.* This is because small adjustments in the parameters that dictate a companies profit margins or demand will not produce as profound behavioural effects as a change from a increasing rate of profit to a decreasing rate of profit. That said, for certain parameters *numerical sensitivity* can still lead to (un)desirable runaway effects, this is, after all the nature of complex chaotic systems. However, the large amount of parameters within these models, not to mention their correlating effects means that an extensive and thorough *parameter sensitivity analysis* is not realistic. Modellers should again use their own judgment to define what parameters they think are behind the largest underlying causes, and work form there.

Behaviour-Mode Sensitivity is the sensitivity of a model in response to particular behavioural scenarios imposed upon it. These can range from altering *fractional coefficients* to change a decaying relationship into one of exponential growth, to forms of *Extreme Condition Tests.* For example, we could set certain stocks to their maximum values and assess the stability of the simulation's response. These calibration methods are good at determining the limits of the model but it is also good to design realistic Behavioural-Modes around real scenarios the system could face. These realistic scenarios should be appropriately labeled and they should be designed based upon the data already collected. The up-shot of creating these scenarios is two fold, firstly it can be used in *Policy Sensitivity Analysis*, and secondly for support in *Systems Improvement Tests.* See (ste, 2009d) for more details on this.

4 Discussion & Further Research

Despite the apparent merits of Systems Dynamics (SD) demonstrated within this report, the limited amount of studies and research in which SD has been applied to operational research is an obvious mark against its credibility.

We have identified two of SDs biggest limitations, its subjective components and the time burden required to perform SD methodology correctly. As we discussed in Section 3.5, there is no way to provide a clear cut and concrete answer on whether we can trust a model or not. The subjectivity inherent in the modelling and validation process can make outside observers feel un-easy about the apparent lack of scientific rigour in Systems

Dynamics modelling. Of course these feelings are valid, the sciences in general, both 'hard' and 'soft' are undergoing a replication crisis (Hudson, 2021). Some of these subjective features are unavoidable, but an actual formalisation of SD methodology within Behavioural Operations such could to improve its credibility. As a textbook Sterman's *Business Dynamics* does this very well, and with the resurgence in Behavioural Operational Research, other attempts are being made to formalise the field(Hämäläinen et al., 2013)(Morrison and Oliva, 2019)(Martin et al., 2016). Examples of rigorous statistical analysis within SD can also help with this.

The other biggest hurdle for SD in BehOR is time and data. The massive work burden required to conduct a high quality study and obtain enough data to validate the results is a big limiting factor in allowing for more wide spread adoption. This fact cannot really be avoided and time will tell if future developments in this field will lead to ways of expediting the process. AI and data automation tools do have potential, but we have not seen attempts to use them in this field.

References

Avi Goldfarb and Botao Yang. Are all managers created equal? *Journal of Marketing research*, 46(5):612–622, 2009.

Erling Moxnes. Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development. *System Dynamics Review*, 16(4):325–348, 2000. ISSN 1099-1727. doi: 10.1002/sdr.201. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sdr.201. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sdr.201.

Julia C. Bendul and Mathias Knollman. The human factor in production planning and control: considering human needs in computer aided decision-support systems. *International Journal of Manufacturing Technology and Management*, 30(5):346–368, 2016. URL https://ideas.repec.org/a/ids/ijmtma/v30y2016i5p346-368.html.

Peter Duchessi and Robert M. O'Keefe. Contrasting successful and unsuccessful expert systems. *European Journal of Operational Research*, 61(1):122–134, 1992. ISSN 0377-2217. doi: https://doi.org/10.1016/0377-2217(92)90274-D. URL https://www.sciencedirect.com/science/article/pii/037722179290274D. IFORS-SPC Conference on Decision Support Systems.

Neil Pollock. When is a work-around? conflict and negotiation in computer systems development. *Science, Technology, & Human Values*, 30(4):496–514, 2005. doi: 10.1177/0162243905276501. URL https://doi.org/10.1177/0162243905276501.

Francesca Gino and Gary Pisano. Toward a theory of behavioral operations. *Manufacturing & Service Operations Management*, 10(4):676–691, 2008. doi: 10.1287/msom.1070.0205. URL https://doi.org/10.1287/msom.1070.0205.

Kunc Martin, Malpass Jonathan, and White Leroy. *Behavioral Operational Research : Theory, Methodology and Practice.* Palgrave Macmillan, 2016. ISBN

9781137535498. URL https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1282088&site=ehost-live&authtype=ip,shib&user=s1523151.

Maria R. Ibanez and Bradley R. Staats. Chapter 4: Behavioral Empirics & Field Experimetns. In Karen Donohue, Elena Katok, and Stephen Leider, editors, *The handbook of behavioral operations*, Wiley series in operations research and management science, pages 121 – 171. John Wiley & Sons, Hoboken, NJ, 2019. ISBN 978-1-119-13832-7 978-1-119-13830-3.

Matthew Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, 83(5):1281–1302, 1993. ISSN 00028282. URL http://www.jstor.org/stable/2117561.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1914185.

Rogelio Oliva and John D. Sterman. Cutting corners and working overtime: Quality erosion in the service industry. *Management Science*, 47(7):894–914, 2001. doi: 10.1287/mnsc.47.7.894.9807. URL https://doi.org/10.1287/mnsc.47.7.894.9807.

Bradley Morrison. The problem with workarounds is that they work: The persistence of resource shortages. *Journal of Operations Management*, 39-40:79–91, 2015. ISSN 0272-6963. doi: https://doi.org/10.1016/j.jom.2015.07.008. URL https://www.sciencedirect.com/science/article/pii/S0272696315000650. System Dynamics: Emergence, Reinforcement, Adaptation and Traps.

John D. Sterman. *Business dynamics: systems thinking and modeling for a complex world*. Irwin/McGraw-Hill, Boston, nachdr. edition, 2009. ISBN 978-0-07-231135-8 978-0-07-238915-9.

Raimo P. Hämäläinen, Jukka Luoma, and Esa Saarinen. On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research*, 228(3):623–634, August 2013. ISSN 0377-2217. doi: 10.1016/j.ejor.2013.02.001. URL https://www.sciencedirect.com/science/article/pii/S0377221713001197.

J. Bradley Morrison and Rogelio Oliva. Chapter 8: Integration of Behavioral & operational Elements Through Systems Dynamics. In Karen Donohue, Elena Katok, and Stephen Leider, editors, *The handbook of behavioral operations*, Wiley series in operations research and management science. John Wiley & Sons, Hoboken, NJ, 2019. ISBN 978-1-119-13832-7 978-1-119-13830-3.

Chapter 5: Causal Loop Diagrams. In *Business dynamics: systems thinking and modeling for a complex world*, pages 137–190. Irwin/McGraw-Hill, Boston, nachdr. edition, 2009a. ISBN 978-0-07-231135-8 978-0-07-238915-9.

Chapter 6: Stocks & Flows. In *Business dynamics: systems thinking and modeling for a complex world*, pages 191–229. Irwin/McGraw-Hill, Boston, nachdr. edition, 2009b. ISBN 978-0-07-231135-8 978-0-07-238915-9.

Chapter 13: Modeling Decision Making. In *Business dynamics: systems thinking and modeling for a complex world*, pages 513–549. Irwin/McGraw-Hill, Boston, nachdr. edition, 2009c. ISBN 978-0-07-231135-8 978-0-07-238915-9.

A. Ziegenbein J. Nienhaus and P. Schoensleben. How human behaviour amplifies the bullwhip effect. a study based on the beer distribution game online. *Production Planning & Control*, 17(6):547–557, 2006. doi: 10.1080/09537280600866587. URL https://doi.org/10.1080/09537280600866587.

Jay W. Forrester. Information sources for modeling the national economy. *Journal of the American Statistical Association*, 75(371):555–566, 1980. ISSN 01621459. URL http://www.jstor.org/stable/2287644.

Chapter 21: Truthe and Beauty: Validation & Model Testing. In *Business dynamics: systems thinking and modeling for a complex world*, pages 845–890. Irwin/McGraw-Hill, Boston, nachdr. edition, 2009d. ISBN 978-0-07-231135-8 978-0-07-238915-9.

Alfred Korzybski. *Science and sanity: an introduction to non-Aristotelian systems and general semantics*. Inst. of General Semantics, Brooklyn, N.Y, 5. ed., 3. print edition, 2005. ISBN 978-0-937298-01-5.

Fabien C. Y. Benureau and Nicolas P. Rougier. Re-run, repeat, reproduce, reuse, replicate: Transforming code into scientific contributions. *Frontiers in Neuroinformatics*, 11, 2018. ISSN 1662-5196. doi: 10.3389/fninf.2017.00069. URL https://www.frontiersin.org/articles/10.3389/fninf.2017.00069.

Parameter Estimation Through Maximum Likelihood and Bootstrapping Methods. In Hazhir Rahmandad and Rogelio Oliva, editors, *Analytical methods for dynamics modelers*, pages 3 – 38. The MIT Press, Cambridge, Massachusetts, 2015. ISBN 978-0-262-33144-9. OCLC: 932293889.

Robert Hudson. Should We Strive to Make Science Bias-Free? A Philosophical Assessment of the Reproducibility Crisis. *Journal for General Philosophy of Science*, 52(3):389–405, September 2021. ISSN 0925-4560, 1572-8587. doi: 10.1007/s10838-020-09548-w. URL https://link.springer.com/10.1007/s10838-020-09548-w.