Reducing the Bullwhip Effect:
Looking Through the Appropriate Lens

Denis R. Towill, Li Zhou*, Stephen M. Disney

Logistics Systems Dynamics Group, Cardiff Business School, Cardiff University,
Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK.
Tel: +44 (0)29 20876915. E-mail: ZhouL@Cardiff.ac.uk

Summary
Demand amplification, now frequently referred to as “bullwhip” is potentially a very costly phenomenon. It can lead to stock-outs, large and expensive capacity utilisation swings, lower quality products, and considerable production/transport on-costs as deliveries are ramped up and down at the whim of the supply chain. However the detection of bullwhip depends on which “lens” is used. This in turn depends on the background and requirements of various “players” within the value stream. To gain insight into this scenario we exploit a relatively simple replenishment model. Because new and novel analytic solutions have been derived for all important performance metrics, comparison of the competing bullwhip measures is thereby greatly streamlined. In the complex real-world the likelihood is that supply chains will generate even greater inconsistency between alternative variance, shock, and filter lens viewpoints.

Key Words: Bullwhip, variance, shock behaviour, resonance, filtering

1. Introduction
Bullwhip is a relatively new phase coined by Lee et al. (1997) to describe the demand amplification phenomenon which was already well known (and reportedly tackled) at Procter and Gamble as long ago as 1919 (Schmenner, 2001). Demand (or variance) amplification of orders as they pass up the supply chain from marketplace to raw materials supplier, and some of the reasons for its existence were also well known to economists some 80 years ago (Mitchell, 1923). Some early OR contributions in minimising cost functions for replenishment systems by Deziel and Eilon (1967) and
Adelson (1966) implicitly covered order variance problems but were not projected in a bullwhip context. Undoubtedly the most seminal contribution to understanding the bullwhip phenomenon was that of Jay Forrester (1958). Based on his extensive knowledge of early computer systems, and supplemented by his understanding of differential equations, he was able to simulate the bullwhip effect on various models and suggest ways of reducing it. The resultant waveform propagation curves then yield “rich pictures” of likely system behaviour (van Aken, 1978).

Forrester’s work is additionally outstanding for two other contributions. He developed DYNAMO, one of the early simulation languages specifically developed for modelling complex dynamic systems. The second, and arguably even more important step, is use of influence diagrams to describe real-world enterprises in an explanatory and communicative manner. This is an ever extending field of investigation as more and more businesses and indeed socio-economic systems are modelled in this way. Notable contributions have been made to this expansion by Roberts (1981), Senge (1990), and Sterman (2000). More recently it has been examined in the context of a product attribute supply chain (Vojak and Suárez-Núñez, 2004).

However despite the undoubted enthusiasm for bullwhip as a research topic, there is a need to exercise caution. A lot depends on the observer and the assumed (or real) operating scenario. Confusion may well occur. Hence benchmark results such as those shown in Table 1 may be quoted entirely out of context. As an example we demonstrate that for a particular system selected to exhibit no “variance” lens bullwhip, there is still significant bullwhip when observed via the “shock” and “filter” lens. The practical outcome is that bullwhip is in reality not a generic term meaning the same thing to all system users. Instead it is application specific. Fortunately any ambiguity can be removed by evaluating proposed solutions in a wide range of simulated operating scenarios. However we do expect carry-over of specific measures enabling bullwhip reduction to be effective in all three domains. The actual impact, however, is a variable between applications.
Table 1. Example of Setting Bullwhip Benchmarks via Simulation of Information Sharing
(Source: Dejonckeere et al. 2004 and cross-checked by Chatfield et al. 2004)

<table>
<thead>
<tr>
<th>Amplification Ratio</th>
<th>Information Mode</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Sharing</td>
<td>With Sharing</td>
<td></td>
</tr>
<tr>
<td>Retailer/Customer</td>
<td>1.67</td>
<td>1.67</td>
<td></td>
</tr>
<tr>
<td>Wholesaler/Retailer</td>
<td>2.99</td>
<td>2.61</td>
<td></td>
</tr>
<tr>
<td>Distributor/Wholesaler</td>
<td>5.72</td>
<td>3.83</td>
<td></td>
</tr>
<tr>
<td>Factory/Distributor</td>
<td>11.43</td>
<td>5.32</td>
<td></td>
</tr>
</tbody>
</table>

2. Bullwhip History

Demand amplification is caused by some internal mechanism or event; it is not due to something external to the system. So although the customer demand may be extremely volatile, it is self-induced worsening of any situation which we are studying here. At least ten causes have been documented (Geary et al. 2003). They may be algorithmic, “player” sourced, or due to poor system design. Furthermore a strong business within a chain can impose a smooth order pattern for the benefit of the pipeline. In contrast a weak business can generate fluctuating orders despite relatively constant demand. Because bullwhip is a time-varying phenomenon, graphical representation of system behaviour is extremely helpful. We now present three examples best viewed through different bullwhip “lens”.

Figure 1. Demand Amplification Typical of Time Series to be Viewed Through the “Variance” Lens (Source: Potter, 2005)
Figure 1 shows bullwhip occurring in the supply chain involving a European fruit juice provider. It is obvious from inspection that bullwhip is present since factory orders are substantially more volatile than the incoming demand pattern. In this particular instance it is clear that a major cause of bullwhip is the batching policy implemented by the factory (Potter, 2005). It is a scheduler response to a problem in “fitting in” the demands imposed by various aggressive customers. As Metters (1997) has indicated, this policy does carry significant on-costs. However this is a situation where policy improvements have been achieved via application of OR statistical type estimation (Potter et al. 2005), hence the relevance of using the “bullwhip variance” lens in this instance.

![Figure 1: Bullwhip Effect in Supply Chain](image)

**Figure 2: Demand Amplification of Time Series to be Viewed Through the “Shock” Lens (Source: Fisher, 1997)**

Figure 2 is a popular example which illustrates how a major manufacturer “gambling” by offering a temporary discount severely disrupts the system both upstream and downstream of the miscreant. This causes a “shock” to the system forcing retailers to stock up rapidly, then run stocks down as they realise customer sales are relatively smooth. “Overkill” is thus followed by “under kill” as warehouses burst at their seams during the run up to Xmas. So space is at a premium and handling costs will
be high. Furthermore the transport companies will also see the same boom-and-bust scenario of working overtime followed by a horrendous slump with drivers laid off. Little wonder that Metters (1997) has shown that total on-costs caused by bullwhip are so high. This situation of “shock” behaviour also has additional problems due to the space dimension (how far away geographically is this earthquake felt?) and the time dimension (when does the wave reach me?). With global supply chains the answers may be thousands of miles and weeks respectively (McCullen and Towill, 2003), furthering the chances of obsolescence, wastage, and bankruptcies. Even when such disruption may be anticipated, the magnitude is unknown, so on-costs inevitably escalate (Walters, 1996).

This shock induced bullwhip might well have resulted from a systems dynamics simulation performed by Sterman (2000), since his objective would be to explain the boundaries to performance. However the usual test input would be the unit step rather than an impulse since this is more realistic for evaluating hardware systems, from which the analogy first came. But Jay Forrester (1958) also detected the importance of “rogue frequencies” via this methodology. In other words, although the input is (say) random, the output is dominated by a regular periodic wave which is clearly self-induced. Figure 3 illustrates the “resonance” problem, a phenomenon well understood by the hardware orientated professionals in automobile suspension and bridge design. In supply chains the effect of such behaviour is to genuinely believe that a seasonal fluctuation exists and try and track it despite the marketplace sales being relatively smooth. Clearly we are interested in viewing such behaviour as a
filtering problem in removing as much “noise” as possible and in order to best identify the actual “message” (Towill and del Vecchio, 1994).

3. Three Bullwhip Lens

OR practitioners typified by Adelson (1966) and Deziel and Eilon (1967) studied relatively simple systems but their methodology required relative complex mathematics for solving the cost minimisation problem. This elegant OR approach is concerned with understanding signal statistics. Hence we term this viewpoint as looking through the **Variance Lens**. In contrast Forrester style modelling and simulation requires a complex model to be built and tested. Until recently this has been a laborious and time consuming exercise. Hence the emphasis on using “dummy” inputs such as step functions which generate “rich picture” responses. The hope and often the reality is that cause-and-effect can be readily identified by just comparing a few graphical outputs (van Ackere et al. 1993). This is most readily achieved if a limited number of key performance indicators are used. A typical set include the peak deviation, the time at which it occurs, and the time then taken for the system to settle back into a quiescent state. Since the idea here is to violently disturb the model, and see what happens or observe violent behaviour in real-world supply chains, we call this perspective the **“Shock” Lens**.

Table 2 “Observer Perspectives and Bullwhip Lens”. Table 2 is available at the end of the document

Table 2 shows how bullwhip observer perspective relates to these lens attributes. But we still need to consider the cause of the many physical and economic systems where periodic behaviour is observed and which is itself regarded as an indicator of “health” or otherwise. In particular the phenomenon of “resonance” in which signals at a particular frequency are greatly magnified is well-known as a characteristic of a poor automobile suspension system or a badly mounted machine tool. Even more spectacular are the resonant effects in large structures such as the Tacoma Bridge which ultimately lead to catastrophe. But this should preferably be seen within the general context of the frequency domain approach. Here the expected real-world excitation signals are notionally considered as the sum of a limited number of discrete frequencies i.e. a Fourier Series representation of the data.
If the system is linear, then these sinusoidal inputs may be decomposed, the outputs to these separately calculated (very easily) via complex number theory, and the total output obtained by summing these individual contributions. Partly as a means of designing complex systems prior to the computer age, and partly because the frequency domain can be related to real-world experience and practice, it has become an extremely popular approach to system synthesis (Towill, 1975). Consequently the frequency domain has even become part of common everyday language (de Bono, 1977). In particular the use of “message” (as a signal to be faithfully transmitted) and “noise” (an unwanted signal to be got rid of) has led to the concept of “filtering” as an action by which these goals are achieved. Hence there is an established and powerful viewpoint of bullwhip as a resonance effect when observed via the “Filter” Lens.

Figure 4. Bullwhip Observed Through the “Variance”, “Shock”, and “Filter” Lens (Source: Authors)
These three bullwhip lens are, of course, related via the system transfer function model as shown diagrammatically in Figure 4. This will be exact if behaviour is linear (Dejonckheere et al. 2002, Disney and Towill, 2002, Disney and Towill 2003, Dejonckheere et al. 2004), but more and more tenuous as non-linearities impact. Nevertheless it is possible to develop “linear” design approaches which will result in acceptable performance in some non-linear systems. In other cases where a system has to operate under a very wide range of conditions, it is helpful to “compartmentalise” the controls so that a different response is called for depending on the instantaneous operating scenario. In turn this necessitates measuring all present system states i.e. “information transparencies” and exploiting this to optimally switch between alternate control regimes. An example of this in the supply chain arena is where a company may decide to move temporarily from “rapid response” mode to “level scheduling” mode as dictated by current market conditions. The Japanese National Bicycle Company is a good example of this strategy (Fisher, 1997).

4. The Demonstrator Model

Our argument here is that we can easily demonstrate the different bullwhip viewpoints via a simple but well understood systems model representing the replenishment schema at just one level in the supply chain. If there can be confusion between the outcomes in this test case, then it is likely that the real-world situation will be even more hazardous. After all there will be much added complexity due to interactions between echelons, effect of product mix, and typical non-linear behaviour due to capacity constraints, stock-outs, and no shows etc. For convenience the model selected and parameter settings adopted are such that all of the bullwhip measures may be obtained analytically. We do not believe this represents a restriction on our approach, especially as our starting point is a well behaved replenishment system. The critical system performance is, however, evaluated by simulation. This not only acts as a cross-check on the methodology, but also reinforces our view that the preferred system design should be tested on a comprehensive but wide ranging set of demands.
APIOBPCS Model Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONS:</td>
<td>Customer Demand Rate</td>
</tr>
<tr>
<td>ORATE:</td>
<td>Order rate placed in pipeline</td>
</tr>
<tr>
<td>COMRATE:</td>
<td>Completion rate</td>
</tr>
<tr>
<td>AINV:</td>
<td>Actual inventory level</td>
</tr>
<tr>
<td>TINV:</td>
<td>Target inventory level</td>
</tr>
<tr>
<td>WIP:</td>
<td>Work-in-process inventory in the pipeline</td>
</tr>
<tr>
<td>DWIP:</td>
<td>Desired work-in-process inventory in the pipeline</td>
</tr>
<tr>
<td>$T_a$:</td>
<td>Time to average consumption, i.e. exponential smoothing parameter</td>
</tr>
<tr>
<td>$T_i$:</td>
<td>Time to adjust WIP error</td>
</tr>
<tr>
<td>$T_w$:</td>
<td>Time to adjust inventory error</td>
</tr>
<tr>
<td>$T_p$:</td>
<td>Actual manufacturing or delivery lead-time days</td>
</tr>
<tr>
<td>$\bar{T_p}$:</td>
<td>Estimated pipeline lead-time days</td>
</tr>
</tbody>
</table>

**Figure 5. The Demonstrator APIOBPCS Replenishment Model Used to Illustrate Bullwhip in the Three Domains (Source: John et al. 1994)**

The replenishment model selected is shown in block diagram format in Figure 5. It is known as the **Automatic Production and Inventory Order Based Production Control System** (APIOBPCS) where to guard against inventory offset, the WIP target is updated in the event of lead time changes, hence $T_{\bar{p}} = T_p$. It is known that APIOBPCS is a generic representation of many practical systems, both algorithmic and human-centred. Sterman (1989) showed that in many cases this model provided a realistic curve fit to decision makers' behaviour at individual echelons in the MIT
Beer Game. A comprehensive transfer function analysis has appeared (John et al. 1994) and exploited in supply chain modelling by Mason-Jones et al. 1997, Dejonckheere et al. 2004 and Chatfield et al. 2004). Hence although arguably simplistic, our selected demonstrator model has the great advantage of common availability.

5. Exploitation of Analytic Solution

| System Transfer Function | \[
\frac{\text{ORATE}}{\text{CONS}} = 1 + s\left(\frac{T_a + T_i + T_p}{1 + sT_a}\right)
\]
| General Variance Ratio Equation | \[
\frac{\sigma_{\text{ORATE}}^2}{\sigma_{\text{CONS}}^2} = \frac{T_a^2 + (T_p + T_i)^2 + Ta(3T_i + 2T_p)}{2T_aT_i(T_a + T_i)}
\]
| Special Case for Unity Variance Ratio* | \[T_a = \frac{2T_p + 3T_i - 2T_i^2 + \sqrt{T_i\sqrt{5T_i - 4T_i^2 + 4T_i^3 + 4T_p + 8T_iT_p + 8T_p^2}}}{2(T_i - 1)}\]
| Step Peak Overshoot | \[SP = 1 + \frac{T_a(T_p + T_a)}{T_i(T_p + T_i)}\left(\frac{T_a}{T_i - T_a}\right) + \left(\frac{T_a(T_p + T_a)}{T_a(T_p - T_i)}\right)\left(\frac{T_i}{T_i - T_a}\right)\]
| Peak Amplitude Ratio | \[Mp = \frac{T_a^4 + 2T_a^3(T_p + T_i) + 2T_a^2T_i(T_p + T_i) + T_a(T_p + T_i)^2 + T_a^2T_p(T_p + 2T_i)}{2\sqrt{T_a^2(T_p + T_a)(T_p + T_i)(T_p + 2T_a + T_i)(T_p + T_a + 2T_i)}}\]

NB. \(T_p = T_p^r\) (i.e. the lead times used in the ordering decision are correct)
and \(T_i = T_w\) (i.e. Deziel and Eilon Settings). * Equation is symmetrical in \(T_i\) and \(T_a\)

Table 3. Bullwhip Analytical Equations Derived for the Demonstrator APIOBPCS Replenishment Model (Source: Authors)

Because the model shown in Figure 5 has been formulated in terms of the Laplace Transform Operator, \(s\), the transfer function is the natural mathematical representation. This enables the system response to any input to be determined. However as the complexity increases so undertaking the necessary algebra becomes tedious and fault ridden. Difficulties multiply when maxima and minima are sought.
However current computer software enables the necessary algebraic manipulation to be accomplished relatively easily. The outputs are the new and novel results listed in Table 3 which also highlights the transfer function of our demonstrator model. It is now straightforward to set a desired value of bullwhip and then establish the combination of parameters $T_i$ and $T_a$ which will satisfy this requirement. One further step calculates the bullwhip metrics $Sp$ (the maximum order rate response to a step input) and $Mp$ (the maximum amplitude ratio in the frequency response) to be calculated.

![Figure 6. “Variance” Bullwhip and “Shock” Bullwhip Contours](Source: Authors)

Our analysis uses the variance measure ($BW$) as a benchmark. In other words if the ratio of (order variance/demand variance) is unity then the argument via the “variance” lens perspective is that no bullwhip occurs. However our analytic formulae enable us to obtain cross-measure graphs such as Figure 6 relating variance bullwhip to a “shock” demand bullwhip measure ($Sp$) for a step input and to the frequency domain directed filter bullwhip measure ($Mp$) as shown in Figure 7. These graphs show that there are wide ranging system parameter settings where the variance bandwidth might be considered optimistic. In other words these are situations which can lead to confusion on the state-of-play in the real world of fast moving products and information flow within supply chains. The vendor (based on peak demand) may therefore wonder why considerable excess capacity is required in his production plant when the operations director insists (using his variance measure) “but there is no
bullwhip present”. So in the next section we select a specific set of parameters based on information contained in these contours to set up a specific bullwhip scenario.

Figure 7. “Variance” Bullwhip and “Filter” Bullwhip Contours (Source: Authors)

6. Testing the Model

To arrive at suitable parameter settings for our test model we make reference to either Figure 6 or Figure 7 to select an appropriate value of $T_i$. Using Figure 6 as an example we see that a unity BW is obtained if $T_i = 2$ days. But to guarantee this result we must solve the $T_a$ equation given in Table 3. Substitution thus yields $T_a = 3.63$ days. Hence the particular demonstrator model parameters are $T_p = 3$ days (exponential lag); $T_i = T_w = 2$ days, and $T_a = 3.63$ days. The experimental conditions now demonstrate the validity of the theory summarised in Table 3 in addition to illustrating the various bullwhip observations.

Figure 8 conceptualises the relative bullwhip measures for this system design predicted to have a unity variance ratio for a random signal demand pattern. In part (a) we show the customer demand and orders placed for this random case. Observation of the ranges suggests that this is indeed a design with little or no bullwhip. But what happens if we assume that demand is suddenly increased step-wise? Surprisingly, at first sight, we see that the orders overshoot by some 50% as visible in part (b). This, of course, is what perturbs the production manager previously told that “there is no
bullwhip”. He suddenly finds that under some conditions he must unexpectedly work overtime, add a shift, outsource, or whatever is necessary to satisfy demand. As Buffa (1969) has emphasised, unit costs thereby increase dramatically, quality may decrease, and there will almost certainly be a learning curve effect as well. Part (c) shows a similar effect in the frequency domain as there is a well defined resonance peak. Not only can this system thereby show self-induced “rogue frequency” tendencies, but similar decision rules in place upstream in the chain can lead to considerable additional volatility.

Figure 8. Comparison of Bullwhip Observed via Three Lens for Conservative System Design
7. Conclusions

Whilst it is true to say that re-engineering the supply chain to reduce any one of the foregoing bullwhip measures means the others are also improved to some extent, this may not occur in proportion. So which bullwhip measure we use should not depend on the mathematics/simulations tools exploited in scheduling packages, but on the user operating scenario. This means looking through the right “lens” (Morecroft, 1983), which for this purpose we shall term as follows:

- Bullwhip “variance” lens
- Bullwhip “shock” lens
- Bullwhip “filter” lens

in response to random demands; “shock” demands; and recurrent peak demands respectively. The first is essentially statistical whereas the last two are much in evidence where “rich pictures” of behaviour are available.

Once it is accepted that there is a problem due to observation via different bullwhip lens, it should not be difficult to resolve it. What needs to be understood is that we are dealing with a complex dynamic system. Hence we may well need a set of performance measures which describe target behaviour under a wide range of conditions. As with hardware systems, supply chains need to cope with many situations analogous to “locking on” (i.e. responding quickly to change), “tracking” (i.e. high availability under quiescent conditions), and “disturbance rejection” (i.e. adequately coping with disruption).

It is not difficult to see that the variance lens concentrates on the tracking model, whereas the shock lens looks at locking-on and disturbance rejection. Arguably the filter lens endeavours via empirical inferences to control all three operating requirements. The best solution is again suggested by considering how designers of hardware systems cope. Their answer is to synthesise in the domain of choice, using a “best practice” database but to cross-check via simulation studies across all possible conditions. But the good designer then learns from each successful system by feeding
back generic lessons into the design process hence keeping the knowledge base growing.

References


<table>
<thead>
<tr>
<th>Observer Perspective</th>
<th>Domain of The Bullwhip Expert</th>
<th>Assumed Demand Pattern</th>
<th>Typical Bullwhip Measurement</th>
<th>Typical Approach Adopted</th>
<th>Systems Performance Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>“Variance” Lens</strong></td>
<td>Operations Researcher</td>
<td>Stochastic Input</td>
<td>Variance ratio, standard deviation ratio</td>
<td>Difference equation with some parameters variable</td>
<td>Minimize cost function</td>
</tr>
<tr>
<td></td>
<td>Pragmatic Manager</td>
<td>(Random sequence maybe obeying a probability distribution)</td>
<td>Extreme swings in order patterns</td>
<td>By direct observation of data to get a good feel for what is causing poor behavior within the real-world “mess”</td>
<td>Simplistic decision making results in either directly Passing-on-Demand or by Level Scheduling</td>
</tr>
<tr>
<td><strong>“Shock” Lens</strong></td>
<td>Control Theorist</td>
<td>Step Input</td>
<td>Peak value of step response; time at which peak occurs</td>
<td>$s$ or $z$ transfer function, and concentrates on system structure, initially to guarantee stability, and then to shape the desired response.</td>
<td>Minimize performance index</td>
</tr>
<tr>
<td></td>
<td>Industrial Dynamist</td>
<td>(Sudden change from one state shift to another)</td>
<td>Peak value of step response; time at which peak occurs</td>
<td>System modeling where causal loop diagrams are transformed into simulation models and studied via test demands</td>
<td>Ensuring we cope with “Shock” demands</td>
</tr>
<tr>
<td><strong>“Filter” Lens</strong></td>
<td>Filter Designer</td>
<td>Frequency</td>
<td>Peak value of frequency response; resonance period; bandwidth</td>
<td>The problem is expressed in frequency domain where value judgments are made on spectrum widths of the “message” and the “noise”, or “disturbances”</td>
<td>Select optimum bandwidth to enable good message transmission plus good noise rejection</td>
</tr>
<tr>
<td></td>
<td>Phenomenologist</td>
<td>(Cycle, Seasonal including rogue seasonality)</td>
<td>Peak value of frequency response; resonance period</td>
<td>Direct Observation of Data</td>
<td>Ensuring we control the resonances and hence avoid rogue seasonality</td>
</tr>
</tbody>
</table>

Table 2. How the Observer Perspective Shapes Bullwhip Perception (Source: Authors)