Big Data Modelling of the Knowledge Economy

Robert B. Mellor
Computing and Mathematics
Kingston University
London, UK
r.mellor@kingston.ac.uk


Abstract

A computer-generated 3D model is used to illustrate the advantages of in silico techniques in virtual reality. The model is derived from theoretical data available for SMEs in service industries and for example enables a business owner (or consultant) to identify where the particular organization is on this three-dimensional landscape and draw quantitative conclusions about fruitful future directions for the organization, as well as how high the benefits will be and what costs will be due along the journey.

Clearly this landscape map is of immense value for academics and practitioners alike, and is available “ready-to-go” in a generic and easily-applicable form. This Masterclass will go further and show participants the use of software tools to construct a contour landscape that is specific for their own client organization.

We will create the basic 3-dimensional fold in Maple and discuss the implications for SME growth and development. We will then move to Markov Chain Monte Carlo modelling, which - put simply - is throwing virtual balls down the basic fold to show how the fold can be used and manipulated to model and predict outcomes of “Knowledge Engineering”. In particular we will model; adding multi-skilled innovators, adding network input from the external environment, looking at the cost of controlling management by showing how much bad information gatekeeping at departmental manager level actually costs and investigating the value-adding effect of intellectual capital.

If time allows, participants will be presented with a small case study and – together with the class leader – will work through a simple solution thus illustrating the ease and applicability of the model in a practical and hands-on fashion.

Keywords: Business Analysis, Computer model, Innovation, profitability, SME.
Introduction

The traditional resource-based view (RBV) of the firm arises from an industrial organization approach that assumes there is a business reality about which information can be gathered and processed to arrive at rational decisions which are subsequently acted upon by management layers in order to realize the strategy involved. The RBV however is very stable and baulks at the concept of fundamental change. Thus a newer approach was born which puts change at the heart of the argument; this Knowledge Based View (KBV) is centred around a firm's dynamic capability which sounds fine but there can be considerable confusion around what that is and how to acquire and develop it, which in turn forms the rich ups-and-downs playground of knowledge management theorists. Some industry areas have emerged in the past 20 years which are more homogenous and thus amenable to modelling, thus the Innovation Based View (IBV) has been postulated (Mellor, 2015a) to be especially relevant for these sectors, meaning that the IBV is an application for some subsets of the KBV. Relevant sectors in particular involve SMEs in service industries because they typically do not have many complicating factors that make the baseline uneven, like foreign subsidiaries, research gaps, patent barriers, factories and raw materials, production lines etc.

As it stands, KBV is about a firm’s aptitude to realize and deal with change, but is very fuzzy about where that change is leading to, and even how to provoke the ‘correct’ change. Thus KBV theory does not deliver a good method to navigate towards success as defined by entrepreneurial standards including firm survival and improved performance in the face of change. However one improvement is that KBV begins to add subjectivity to strategy by admitting that owners and managers can make mistakes and indeed the term “core rigidity” is now almost as common as the expression “core competencies”. None the less, innovation (in the sense of KBV) management research reveals its deficiencies by still being largely unable to explain precisely how Schumpeterian (i.e. transient) competitive advantage comes about, and how this entrepreneurial effect could be better sustained.

Improved scientific methods, including control experiments, are needed to pin down this important question but these techniques are rarely applicable in the fields of Economics and Business Studies; indeed where such instances do occur they are serendipitous and often oblique. This problem has led to an innovative laboratory approach being developed in computer-generated realities and has the advantage that experiments in silico are not only more rigorous but also cheaper and faster than conventional approaches e.g. case studies. One such computer model has resulted in a data-driven 3D model which completely explains the phases reported 40 years ago that are characteristic of the development of SMEs (e.g. Grainer, 1972). The computer model has also been used experimentally to predict outcomes associated with structural changes within organizations, and preliminary results are encouraging (Mellor 2014a, 2014b and 2015b). The present model is derived from theoretical data (Mellor, 2011) and for example enables a business owner or consultant to
identify where the particular organization is on this 3D landscape and draw quantitative conclusions about fruitful future directions for the organization, how high the benefits will be and what costs will be due along the journey.

To test the hypotheses involved and to perfect this model we have chosen a “Big Data” approach using statistically significant amounts of real data from a longitudinal dataset housed at the Leibniz-Institut für Wirtschaftsforschung (Halle Institute for Economic Research) in Germany. Using the German data in the UK computer model we aim to present a rigorous and evidence-based unified model of business growth and development.

This communication details the model and furthermore provides twelve practical pointers of use to academics and practitioners alike.

Methodology

SMEs in service areas are normally unencumbered by such assets as production lines, patents etc and compete by incremental innovation, a process of being just ahead of the rest in some small way. This “mutual inspiration” is derived in many ways, often by communication between individual of different backgrounds, or outlooks, or educational specialization. Thus diversity adds to innovation and this “Diversity Innovation” is thought to be the most important form of innovation in small service industries because it provokes a valuable “mutual inspiration” effect. To quantify this effect persons are represented as nodes (the number of people is represented by P) and they are joined by ties, the number of links or ties between nodes is the Diversity Innovation (DI) number. As the DI number increases the potential for innovation increases. Clearly the DI number merely represents the potential for the type of small-scale, incremental innovations thought of as being important in small growing organizations (SMEs) and not the sum knowledge in the organization. A simple equation has been published (Mellor 2011) that shows that the larger the number of people in communication, the larger the probability that some kind of “mutual inspiration” will occur. Clearly one could be tempted to put lots of different people in a room and wait for an innovation to take form but in mercantile organizations the transaction costs for this type of communication are too high. To illustrate this, think of a company with 120 employees; the DI number = 7140, so talking continuously and without any break, this would take 595 hours or 16 man-weeks of working time just for employees to talk to each other for 5 minutes, excluding that any employees got a chance to repeat any conversations (or do any work). It would be expected that the total outlay even only in terms of salary for this “speed-dating” exercise would exceed the value of any innovation arising from it. Even worse, each further employee hired would take around 10 man-hours to talk to existing employees for 5 minutes each, so the transaction costs are considerable.
There are also transaction costs involved in management, firstly that as departments grow to size 50 employees, the transaction costs incurred by the manager mean that the manager uses all their available time to communicate with them and that more employees means too little time and a consequent reduction in communication quality, thus organizations tend to enter into a policy of serial departmentalization during growth when any department reaches maximum 50 employees. As the organization grows, it is split into departments and this has consequences for the DI number, as shown in figure 1 (below).

**Figure 1**: DI number (subjected to a smoothing algorithm) in a population between 1 and 250 nodes (employees), splitting the organization into equally-sized parts when each department reaches size 60.

To this 2D curve we can now add the third dimension. The third dimension of the "landscape" is given by the J-curve (a hockey stick curve), because the J-curve represents a transition where things get worse before they get better, this effect is known from several disciplines, e.g. medicine, national economics and Business Process Re-engineering. J-curves are known from national economic data (shown in figure 2 in blue) but in this case data calculated from the Business Process Reengineering literature is used (violet stippled in figure 2). The difference between these curves is that bankrupt states can keep going, but bankrupt companies (in a Capitalist democracy) cannot.
Figure 2: The financial journey from low (left) to high (right) value for national economic development (blue line) or companies undergoing re-organization (stippled line).

On this axis there are no units for innovation ("innos"). However we can use % instead on a 1-100% scale where the value is the amount derived from the maximum (100%) possible. We can do this because the maximum possible, for an organization of known size (number of employees) is known from the DI number. A finer granularity can be obtained by benchmarking the degree of innovation the firm involved against other firms in the same industry segment.

When the curve (for the business process reengineering data) from figure 2 is added as the third dimension to figure 1 using Maple software (maplesoft.com), then the following 3D fold is obtained:

Figure 3: Figure 3A, the basic fold and figure 3B illustrating the axis dependencies; X axis as growth (number of employees), Y axis as value (e.g. annual turnover) and Z axis as openness to innovation on a benchmarked scale.
Results from modelling in Maple.

From figure 3 any business owner (or consultant) can identify where the particular client organization is on this 3D landscape and draw quantitative conclusions about fruitful future directions for the organization. However the fold in Figure 3 does not indicate the costs associated with any future plans. To estimate costs figure 2 is revisited, because figure 2 shows that the J-curve slips below zero. In the analogy (to a valley) presented here the zero level is thought of as being represented by a lake (the ‘tarn’ of ‘knowledge valley’). Organizations progressing from a low-innovation state to a high-innovation state, for example by the process known as Business Process Reengineering, often transiently lose profitability. Using figures from the relevant literature, Mellor (2011) derived two important factors:

1. That for each individual firm the drop in profitability can be greater or lesser, but the maximum drop is around 40%, greater loss than this generally leads to rapid downsizing and the shedding of knowledge assets and core competencies which in turn leads to bankruptcy.

2. The time of transition is important, the shorter the time in the loss-making zone the lesser the chance of negative consequences like bankruptcy.

The ‘tarn’ fits into the fold as shown in Figure 4.

**Figure 4**: The fold up to organizations of size 200 employees showing the position of the zero profitability-level as a lake. Using the terminology of small business research, companies just starting up are designated ‘mice’, those at the peak of performance are ‘gazelles’ and less innovative medium-sized enterprises are ‘elephants’.
Figure 4 indicates two further general lessons to ponder upon:

3. The second round of departmentalization cuts profitability back to near-zero levels. Clearly while owners are growing their organization family or friends may be installed as managers, but at the second round this becomes a dangerous practice and consultants etc should be recommending that competent professionals are hired.

4. The depth of the ‘tarn’ increases with size, so the costs for a medium-sized organization to cross to the innovative side are more than the costs for a small organization. So organizations that from the very beginning stay on the innovative side avoid costs and danger later in their life.

Figure 4 also illustrates that around the zero level (on both sides) is the home of social entrepreneurship, those organizations that aim at rarely or only transiently making a profit and these may be churches, charities or similar. Organizations can also exist beneath the zero level; these are for example state organizations. These adopt a pendulum path from side to side, rarely surfacing above the zero level. For example state bureaux exist in the depths of the tarn because the pull up to the right of the J-curve towards efficiency and better public service (modernization & e-government initiatives, etc) is balanced by a pull to autocracy and the left of the J-curve (political expedients, the dominance of political targets, civil service “empire-building”, etc), leaving them always loss-making at around the lowest point.

Furthermore if any state service does become (potentially) profit making then state policies may well indicate that this portion is to be sold off to private enterprise, leaving the loss-making remainder to sink down again to the mid-point.

It is known that unhindered knowledge flow and the rapid dissemination of incremental innovations arising from this, is especially important in SMEs (Desouza & Awazu, 2006. Serenko et al, 2007) and this aspect has been applied to studies on management control in SMEs (for some examples see e.g. Ditillo, 2004). In terms of management control, Williamson (1985) argues that "self-interest seeking with guile" (Williamson, 1985 p30, Williamson 1993, p97) is inherent in human behaviour and thus that it is difficult to identify trustworthy individuals in organizations. The logical result of this is that it is necessary for organizations to structure themselves with internal checks and balances in such a way as if individuals cannot be trusted, which clearly adds a further layer to the internal transaction costs that organizations incur.

While this is very rational, it begs the question; how high is the cost of hindering knowledge flow at departmental manager level? This is because it is inefficient to have expensive policing if the cost of the crime is petty. Modelling can give an indication of the cost because the links between departments can be discounted in the algorithm. Figure 5 shows the effect of hindering knowledge flow at departmental manager level.
Figure 5: The fold with; figure 5A, open information gateways at departmental manager level and figure 5B, closed information gateways at departmental manager level. Note in addition to poorer performance overall in figure 5B, there is a further performance dip (indicated) in figure 5B.

With all four inefficient department leaders and the organization at the peak of performance, the point where the most difference would be found, overall performance was down by a mere 6.4% (SD 0.05, n=10). As an across the board rule and subject to some variation (within the constraints given in real life like the vagrancies of exactly how poor the information gatekeeping is) then this (6.4%) percentage of financial under-performance will apply to organizations generally having four departments. Further modelling showed that one departmental manager blocking information and knowledge flow to other departments reduced the financial performance of an organization by 1.4% in the local department and a further 1.2% spread across the remainder of the organization. Two such managers reduced overall performance by 4.1%. Results indicate:

5. In high-innovation environments the costs of "self-interest seeking with guile" are not sufficiently high to justify heavy preventative measures, at least in the short-term. However, if the workforce attached to that departmental manager becomes demotivated – a scenario that may be more common in low-innovation work environments – then costs may increase significantly.

6. In the “closed gate” scenario a crisis occurred at the third round of departmentalization at around size 150 employees which is not seen in open gatekeeping. Perhaps this indicates e.g. the departmental managers squabbling over the division of budget and assets as the organization grows from three departments to four departments - costly disagreements.
Thus in addition to locating where an organization is on the knowledge valley fold (figure 4) one can now draw conclusions about the amount of resources that should be deployed to ensure the internal policing of the departmental managers behaviour.

**Results from Markov Chain Monte Carlo modelling.**

Although the fold shown in figure 3A is not a true Markov Chain, it can be used as such for the purposes of modelling. Markov Chain Monte Carlo simulations (see e.g. Chib and Greenberg, 1996 for a classical review of Markov Chain Monte Carlo techniques or Robinson, 2014, for a more recent overview) in this case start by rolling virtual balls from the ‘mouse’ origin down the fold. Everything being equal, the balls will hit the far end of the J-curve in a random fashion i.e. there will be a homogenous distribution of exit impacts along the curve. However the fold can also be manipulated e.g. squeezed or attenuated, and the change in distribution of the exit impacts visualized as a scatter plot. If a statistically significant correlation exists between the factor being tested and adding value, then the scatter plot will form a probability density function (a Gaussian distribution or ‘normal curve’). On a classical Gauss curve the scale parameter (small sigma, $\sigma$) is 1.00. How much a given curve varies from a classical Gaussian distribution can therefore be assessed by measuring the scale parameter of the curve; the more it diverges from 1.00 the less correlation there is. Values of scale parameter below 1.00 indicate a very sharp peak and those over 1 indicate a spread peak (referred to as a platykurtic distribution), i.e. of low correlation. Values of 3.0 and over indicate a correlation that, for practical purposes, is so spread and weak that it can be ignored.

Workplace innovation/mutual inspiration is at its most simple - and powerful - when communication problems do not exist, i.e. where the two or more specialists are literally embodied within the same person. Thus one person – one human brain – is able to look at a problem with the eyes of (say) a geologist and a businessperson, or teacher, simultaneously. Indeed employees with a multi-specialised background i.e. possessing ‘T-shaped’ or ‘A-shaped’ skills (Tsai and Huang, 2008; Mellor, 2005) have been implicated in raised innovation. Furthermore they are likely to be highly qualified and thus may be assumed to work in middle management. Accordingly the second dimension (value) in the knowledge valley fold can be converted from company annual turnover to salary, and a ‘knowledge trail’ (Mellor, 2011) spanning a reasonably wide band of middle management salary inserted. This band thus forms a bifurcating pathway along the fold and exits occur on both the left and right-hand sides of the fold on the Z-axis. Along this bifurcating band, innovators are represented not as one node, but as two overlapping nodes, i.e., that no tie is needed to traverse between the nodes. This is done to represent the lack of transaction costs for communication in this individual, so in this case the Monte Carlo ball can progress two nodes down the fold without incurring ‘costs’. The results obtained are shown in figure 6 and illustrate that low-innovation companies i.e. on
the left side of the J-curve) appear to be able to profitably use multi-specialised individuals in middle management and indeed things may seem to get quite a lot better (remember the vertical axis is value) as the organisation opens up to innovation and change. However, as they proceed along the skewed Gauss curve (moving right) they approach a decline, the nature of which could be very diverse depending on branch and industry, but in any case may be disadvantageous.

Figure 6: Showing (Figure 6A) the relationship between placing innovators in a band marked by the block, and exit impacts on the J-curve. Figures 6B and 6C; scatter plot values derived from exit impacts on the both sides of the fold.

Figure 6 also shows that companies showing a high percent of use of innovation can likewise profitably use multi-specialised individuals in middle management. The model indicates that only low or no risk is associated with this strategy and, if it works, things may get much better very rapidly indeed. After this gain in value is realised, added value will reach a plateau and level off and thereafter increasing the numbers/density of multi-skilled individuals (i.e. hiring more) should not have much further effect. Thus:

7. Adding innovations/innovators to low-innovation organizations can have positive effects relatively quickly (the “low-hanging fruit” effect) but the temptation to add more and more innovation should be resisted unless a full-scale business process reengineering operation is planned.

8. Adding innovations/innovators to high-innovation organizations is low-risk and can have some positive effects but these are more likely to be reaching performance gains more rapidly than overall performance improvement. One way in which organizations can profit would be in cases where these innovations are poorly aligned with ‘core competencies’ and the parent organization can profitably spin-out and later sell these innovations encapsulated in a start-up firm.
In the model used above, innovators were added into the bifurcated stream at random. Clearly models can be produced where they are located more upstream or more downstream. A number of innovators were used where the scale parameter (small sigma, \(\sigma\)) was 1.5 and the experiment repeated with a Pareto distribution upstream or downstream, the results are shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>80% upstream, 20% downstream</th>
<th>Random (control)</th>
<th>20% upstream, 80% downstream</th>
</tr>
</thead>
<tbody>
<tr>
<td>High innovation side</td>
<td>1.21</td>
<td>1.50</td>
<td>1.61</td>
</tr>
<tr>
<td>Low innovation side</td>
<td>1.22</td>
<td>1.50</td>
<td>1.63</td>
</tr>
</tbody>
</table>

**Table 1**: The effect of the placement of innovators on the value of the scale parameter (\(\sigma\)) in the resulting scatter plots (SD 0.05, n=10).

Table 1 shows that in both high- and low-innovation firms, a ‘tighter’ peak was obtained when the innovators were placed upstream. This implies:

9. Putting innovations and innovators into an expanding organization is good, but the effect of parachuting innovators onto an already existing and less innovative middle management, is less effective than having started with an innovative middle management right from the beginning.

In the modelling described above, a rather high number of innovators has been used; to achieve a scale parameter (\(\sigma\)) of 1.5 in the scatter plots the number was 50% of all middle managers. Clearly this is rather a lot and at odds with everyday reality. However innovations do not have to have an endogenous source i.e. arise *in situ* from within the firm, they can arrive from outside the firm and be imported. To do this one typically needs a network. Networks can be modelled, in this case Markov Chain Monte Carlo modelling was performed as before but previously all Monte Carlo balls started at the origin (the ‘mouse’ position in figure 4) of the fold and a variable number of ‘innovators’ were placed in a band corresponding to a middle-management salary as above. However in the experiments described in this section packets of useful information could arrive from individuals outside of the organisation. Thus in this experiment the same overall number of balls were rolled down the knowledge valley fold as before, except that a variable proportion of these balls appeared ‘spontaneously’ in a random fashion within the innovator band. For example, 50:50 represents a situation where each innovator can get 1 datum of “just-in-time” knowledge from his/her external network. Experiments tested the proportion of balls at 33: 66 and at 50: 50 and again at 66: 33. This is meant to simulate the
number of innovations (number of balls) being constant while the number of innovators responsible for them was varied, the balls appearing at random represent inspiration coming in from outside the organisation (and thus can appear anywhere along the band). The ratios represent one innovator using their network to harvest two innovations (33.3: 66.6), one innovator bringing in one inspiration from outside (50: 50) and finally the network value being one inspiration from outside for every two innovators in the organisation (66.6: 33.3). The results are shown in table 2:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Value of scale parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>One innovator and two randomly-arising external innovations.</td>
<td>1.15</td>
</tr>
<tr>
<td>One innovator and one randomly-arising external innovation.</td>
<td>1.41</td>
</tr>
<tr>
<td>Two innovators and one randomly-arising external innovation.</td>
<td>1.49</td>
</tr>
<tr>
<td>Innovators only (control)</td>
<td>1.50</td>
</tr>
</tbody>
</table>

**Table 2:** The effect of changing the proportion of Monte Carlo balls starting at the origin of the fold, on the scale parameter of the resulting scatter plot (SD 0.05, n=10).

The results of these experiments show:

10. The ratio of innovators to non-innovators in management can be as low as 1:6 if working "just-in-time" knowledge networks are available.

11. The value of exogenous innovation when arriving in a timely and applicable fashion appears to be only marginally less valuable than ‘home-made’ endogenous innovation arising *in situ* in the organization from multi-skilling itself, it may even be able to compensate somewhat for non-innovative management, always providing that the internal information gatekeeping is both open and open to accepting change.

Please note that the positive effects of exogenous networks does not immediately translate into having large exogenous networks, the innovation arriving has to have immediate applicability, implying that external networks need to be highly relevant rather than simply extensive.

**Intellectual Property and the concept of knowledge trails.**

For the past thirty years the concept of “relentless innovation” has gained wide currency. Despite this and especially in the SME sector, the majority of companies
are reluctant learners, happy to exist, to be making a profit, going for moderately safe goals in inherently unstable Porterian “weather” and unwilling to disturb the balance with risky changes. This is especially true for the majority as they age, become stable and thus become “the incumbents”. They absorb data from the environment and process it according to pre-set routines. These routines can be represented in simple terms as a “preferred relationships” or a Knowledge Trail.

Knowledge trails do not themselves represent an innovation, but they do represent a cost-saving and they are therefore accorded value, the extent of which is dependent upon where they are – high or low – in the knowledge valley fold they inhabit. A mundane knowledge trail could be e.g. customer complaints, where any complaint is processed according to a well-known procedure and employees do not start from scratch each time a customer communication arrives.

Knowledge trails are also associated with intellectual property. A relatively high value knowledge trail could be e.g. the fairly routine registration of trade-marks, design or database rights etc. Clearly of even higher value would be an example concerned with the invention of a new process or device then writing and filing for patent protection. Interestingly the value and trajectory of a knowledge trail is fairly constant, even when the value of the company falls. This is illustrated in figure 7.

![Diagram](image.png)

**Figure 7:** The knowledge valley fold showing (red) a meandering low-level knowledge trail and in contrast a knowledge trail (blue) that is facilitated by the “bridge” of a patent (or other intellectual property).

Thus it becomes apparent:

12. Patents (and proportionally other intellectual property) enable knowledge trails to firstly retain value but also to bridge any gap in organizational profitability during growth which in turn enables the respective knowledge trail to start again at an over-proportionally (compared to the organization) high valuation as support returns.
This explains the popularity of patenting and intellectual property amongst aspiring high-innovation start-ups in e.g. biotechnology because it retains value also when the organization may otherwise be struggling.

**Conclusion**

One should not confuse conclusions drawn from observation with those derived from scientific experimentation. Less than ten years of *in silico* experimentation has explained rationally countless observations made in the field for over fifty years. Not only are these techniques one or two orders of magnitude faster than conventional methods but they are also rather simple to apply. Throughout this communication twelve points have been made which the practitioner can easily appreciate. Academics and practitioners alike can now pinpoint their clients’ position on the knowledge valley fold and for the first time are able to advise clients in quantitative terms the best way forward and pitfalls to avoid.

**References**


