The Achilles' heel of scale service design in social security administration: The case of the United Kingdom's Universal Credit

Professor John Seddon and Brendan O'Donovan

Authors:

John Seddon is the leader of the Vanguard organisations, providing consulting services around the world. He has received numerous academic honours for his contribution to management science.

Brendan O'Donovan is the head of research at Vanguard UK.

Keywords: Universal Credit, economies of scale, digital by default, Vanguard Method

Abstract

This paper takes a critical view of the UK government's design for the delivery of the 'Universal Credit' (UC) benefit reforms. It is argued that the UC is destined to fail because of the policy's extension into specifying the means ('digital by default') of delivery for such services. The authors argue that an unseen but ubiquitous set of 'scale' management assumptions has been allowed to infiltrate the means by which the government intends to enact its headline policy objective to 'make work pay'. Following Seddon's 'Vanguard Method', a practical example of how a better service was designed in a local authority housing benefits service is then examined. Results from this service include being able to deal with up to 50% more demand, with fewer resources, in half the official target time. Finally, the paper will conclude with a call for more evidence-based policy.

The flaws of scale assumptions

Locally designed services are human, receptive, engaging and productive. Counter-intuitively, they are also high quality and low cost. Local authorities that have rejected the government's 'best practice' guidance on how to manage housing benefits now deliver a service that puts official targets in the shade and cuts costs into the bargain. East Devon and Stroud councils, to cite just two, process benefits in less than half the official target time, in a period when the number of claimants increased (Middleton 2010). East Devon serviced 33 per cent more demand and Stroud 50 per cent more, in both cases using less resource. Blaenau Gwent leapt from the bottom of the Welsh league table to the top; the improvement in housing benefits service cut the number of 'benefits' calls to the service centre by 50 per cent and face-to-face visits to solve problems by 57 per cent (Zokaei et al 2010). These were improvements and savings that would never have been put in a central government 'plan'.

However, there is a difficulty. Unfortunately, the idea of locally designed, human-scale service runs against the grain of successive UK governments' policies for public sector reform – in particular, the coalition government's plan to reform the social security benefits system by the introduction of the new 'Universal Credit' (DWP 2010). The authors support the government's stated intention to redesign social security benefits in order to 'make work

pay' (Daily Telegraph 2011b). However, ministers have also stipulated policy which dictates how the Universal Credit must be delivered. The government wishes to use the welfare changes to further its intention to make all public services 'digital by default' (Cabinet Office 2010). Claimants will be pushed towards an online delivery channel and, if that fails to meet their needs, a centralised call centre. As a corollary, the government intends to close down locally-based housing benefits offices.

Examination of what actually happens in IT-dominated industrial designs such as the one proposed for Universal Credit reveals massive disruptions to the service flow, for the customer service is anything but smooth. There is huge waste in the shape of handoffs, rework, duplication of effort, and a focus on meeting activity targets and service levels. All of these 'system conditions' lengthen service delivery times and consequently create failure demand: '*demand caused by a failure to do something or do something right for the customer*' (Seddon 2003 p26). In other words, the service gets worse and costs go up. On all counts such industrial 'scale' designs fail miserably.

Before investigating the design flaws of these proposed welfare reforms in greater detail, it is necessary to briefly outline why industrialised, alienating service designs for the delivery of social security programmes have continued to prove so appealing to politicians. The history of designing work operations to achieve economies of scale is a long one which has strongly influenced managers and politicians alike.

Scale thinking: the milestones in operations management history

Economy of scale, in common with many modern management principles, derives from economics. When more units of a good or service can be produced on a larger scale, with (on average) lower input costs, economies of scale are said to be achieved. The development of the idea can be traced back to the father of modern economics, Adam Smith. On the back of a British £20 note there is a picture of Smith and the words:

> 'The division of labour in pin manufacturing: (and the great increase in the quantity of work that results)'

In his most famous work, 'The Wealth of Nations' (Smith 1998, first published 1776), Smith explained how while one worker could make 20 pins a day, 10 workers dividing the 18 pin-making steps between them could produce 48,000. He identified the division of labour and specialisation as the two key means to achieve this huge increase in productivity. Through these techniques, he believed that employees would not only be able to concentrate on a specific task, but with time, improve the skills necessary to perform their jobs, which could then be carried out even better and faster. Hence, through such efficiency, time and money could be saved while production levels increased: the idea that economy could be achieved through 'scale' production had been established. It is pertinent to note that Smith's innovation also relied on mechanisation, something peculiar to manufacturing systems.

The next boost to scale thinking came with the early 20th century pioneers Frederick Winslow Taylor and Henry Ford, who devoted their efforts to boosting productivity through the specialisation and standardisation of work. F W Taylor developed what became known as 'scientific management' (Taylor 1998, first published 1911). He brought the notion of 'method' to management (he called it the search for 'the one best way' of completing a task), and this spawned 'organisation and methods' departments in every large organisation.

'Method' was established as belonging in the province of supervision: managers saw it as their job to decide the best way to carry out a task and it was the workers' job to follow the specified method.

Perhaps the single most important contribution to the dominance of scale in management thinking, however, was Henry Ford's mass production system (Ford 2003). In the early 20th century, his black 'Model T' cars flowed out of a factory that worked like a grand machine, with men and materials in harmonious flow. His innovation enabled him to halve the costs of production and double the workers' wages. It caught the world's attention; the mass-production factory had brought efficiency, and efficiency meant the ability to compete. Ford's plants were envied throughout the world for their scale and efficiency (Krafcik 1988). Mass production, at scale, became the norm. Inevitably, Ford's methods were copied by other companies, and management writers began to detail the application of Taylor and Ford's approaches in other organisations: 'factory management' was born (Lockyer 1962). The field was gradually extended to become 'operations management' in the 1970s, expanding until it encompassed service operations as well as manufacturing (Johnston 2005).

Scale thinking and the move towards industrialised, standardised service

In 1972, Ted Levitt wrote a seminal Harvard Business Review (HBR) article entitled 'Production-line approach to service'. In it, Levitt urged managers to pay the same attention to improving the efficiency and design of services as they did to manufacturing operations:

> 'In sum, to improve the quality and efficiency of service, companies must apply the kind of technocratic thinking which in other fields has replaced the high-cost and erratic elegance of the artisan with the low-cost, predictable munificence of the manufacturer.' (Levitt 1972 pp43-44)

Levitt used the example of fast-food production and service in McDonald's as an example of how factory methods could be profitably employed in a service. The method by which McDonald's achieved its market domination was through mastery of a 'system' which was 'engineered and executed according to a tight technological discipline that ensures fast, clean, reliable service in an atmosphere that gives the modestly paid employees a sense of pride and dignity' (p45). Levitt believed that McDonald's had successfully applied 'a manufacturing style of thinking to a people-intensive service situation' (p45), reinforcing the argument that service design should employ the manufacturing approaches of standardisation and specialisation.

The next important step in service industrialisation was taken by Richard Chase, who in another HBR article proposed that service operations should be organised in separate 'front' and 'back' offices (Chase 1978). In essence, his argument for a separate back office was that since it has no contact with the customer, it offers greater potential to operate at peak efficiency. Chase argued that service systems with high customer contact were more difficult to control and rationalise than low contact systems; so decoupling front from back enabled what he saw as the 'technical core' to operate as a factory, isolated from outside influences, following a resource-orientated schedule and optimising efficiency through batch scheduling, forecasting, inventory control and work measurement, just as in manufacturing.

These ideas continue to form the conceptual foundations for the way services are designed and managed today. As the American economist-turned-management-guru Michael Porter

puts it:

> 'Scale economies can be present in nearly every function of a business, including manufacturing, purchasing, research and development, marketing, service network, sales force utilization, and distribution.' (Porter 1980 p7)

The 'scale' archetype for service design

The 'scale' archetype for service design has thus emerged from several generations of conventional management theory. Many of the underlying assumptions are so deeply pervasive in management literature and practice that they have become invisible – hidden in plain view. The 'scale' design archetype commonly features:

- Standardisation and specialisation

- Separated and 'optimised' front and back offices

- Access channel management (pushing customers to transact with the organisation through cheaper channels such as by phone or online).

Following this archetype, management is primarily concerned with managing activity on the assumption that activity represents cost. To that end, 'command and control' management is focused on what Seddon describes as the 'core management paradigm' (Seddon 2008 p51):

- How much work is coming in?

- How many people do I have?

- How long do people take to do things?

However, studying service organisations reveals that industrialised designs have an unexpected Achilles' heel: paradoxically, attempts to manage costs create costs.

As a simple illustration, consider what happened when English local authorities were set a target to establish call centres by April 2005. When consultants were hired to help them move 'telephone work' from council departments to centralised call centres, call volumes shot up. Why? The increase in call volumes was caused by the phenomenon of 'failure demand' (Seddon 2003 p26, see definition above). The assumption that telephone work could be treated as a specialised activity separate from the core service provision was an example of misplaced faith in scale and centralisation; as a direct result, call centres were besieged by people wanting to know what had happened to their application or enquiry. Installing more IT in the shape of 'customer relationship management' (CRM) systems only served to institutionalise this waste, compounding the error.

Conventional managers have been taught to strive for lower unit costs, which in a service environment leads them to focus on preferred (cheaper) modes of transactions with their customers. Unfortunately, their focus on transaction costs blinds them to the fact that, while transaction costs may indeed be lower, transaction *volumes* will inexorably rise as the system fails to provide a service that works for the customer. A common reaction to increasing call volumes (as in the local authority call centres described above) is to add more resource – employ more agents to field the calls – which of course mitigates or nullifies any gain from lower transaction costs. Another is to outsource call-handling to lower-wage economies, ignoring the fact that such contracts are commonly priced according to transaction volumes.

Being paid by volume, outsource providers have no incentive to tackle and remove failure demand. Indeed, suppliers are effectively incentivised to worsen service to the customer.

In the same pursuit of lower transaction costs, managers focus on 'access channel management', driving customers away from supposedly expensive face-to-face contacts to the telephone; and then from telephones to electronic transactions online. Such scale concepts are aggressively marketed by the major management consultancies which have developed lucrative businesses providing the necessary IT. Government and the consultancies enjoy a close relationship, with many of the top public sector chief information officers having previously worked for these IT suppliers (Computerweekly.com 2012)

But scale fails: the problems revealed when managers study their services as a system

The Vanguard Method was developed to help managers look at their organisation as a system and, on the basis of the knowledge gained, redesign their services to improve performance and drive out costs. Studying the organisation as a system reveals counter-intuitive truths, the importance of demand, and, in particular, failure demand and its causes, being just one. By studying demand, managers are able to understand the levels of failure demand they are unwittingly creating for themselves and other downstream organisations. Rising failure demand is a signal of a broken system. Revealingly, the 'system conditions' (Seddon 2003 p112) which cause failure demand are found to be those listed above as the ingredients for efficiency through scale. One of the exemplars of this scale design is Her Majesty's Revenue and Customs (HMRC), whose failure to provide a service to many taxpayers is discussed later in this article.

Despite being offered abundant and growing evidence of the costs of scale initiatives, successive UK governments have continued to be seduced by the potential efficiency gains scale thinking promises.

UK government reforms and the growing influence of scale thinking

In the 1998 Comprehensive Spending Review (CSR), the then Labour government promised to 'root out waste and inefficiency' by setting efficiency targets for key public services (HM Treasury, 1998). At the same time, Prime Minister Tony Blair was setting a target for all government services to be accessible online by 2008, later brought forward to 2005 (Cabinet Office 1999). By 2004 and the next CSR, Sir Peter Gershon was employed to find £21.5bn of public sector efficiency savings (Gershon 2004). Among other things, he offered up greater use of IT in benefits payments as an example of 'streamlining the delivery of services to the public, cutting transaction costs and reducing paperwork' (Gershon 2004 p1). Official 'e-efficiency' reports then started to come thick and fast. In 2005, 'Transformational Government' (Cabinet Office 2005) pinpointed the need to 'Develop modern channels for citizen and business access to services, and actively manage the shift in channels towards the most efficient and effective.' In 2006, the Varney review asserted that 'Technology has enabled a revolution in the way service providers interact with their customers' (Varney 2006 p1) before going on to recommend that:

> '... this approach could drive out [in]efficiencies by improved performance and coordination of front-line e-services, contact centres and local offices and reducing duplication of business processes through shared use of an identity management system. Over the longer term further efficiencies and service enhancements could

be made by reducing the back-office functions that would no longer be required'. (Varney 2006 p4)

After the formation of the Conservative-Liberal Democrat coalition government in 2010, there was even greater pressure for public sector efficiency gains in light of the government's stated central aim of reducing the fiscal deficit. The internet entrepreneur Martha Lane Fox was commissioned to write a report reviewing the government's web-portal service, Directgov. Among her recommendations was that government services should be 'digital by default':

> 'Go digital only: ... shifting to digital-only services has huge cost-saving potential. Directgov should be the default platform for information and transactional services, enabling all government transactions to be carried out via digital channels by 2015. Achieving this will require a radical reallocation of effort and resources within Directgov. The organisation must focus its effort on creating high-quality user-friendly transactions and guidance. It can only do so by scaling back on non-core activities and being given the power to enforce user-centred quality standards across government.' (Lane Fox 2010)

Her proposal was enthusiastically endorsed by ministers:

> 'Public services should be delivered online or by other digital means, the Government has announced in response to a report published by Martha Lane Fox today. The report, and the Government's initial response, argues for a Channel Shift that will increasingly see public services provided digitally 'by default'.' (Cabinet Office 2010)

Behind the 'digital by default' policy is the belief that overall costs can be controlled by reducing the cost per transaction. But if going digital increases the total number of transactions it takes to receive service, the cost advantage will be nullified. Our contention is that 'digital by default' will lead to rising failure (and thus overall) demand at a systemic level, substantially increasing costs as a result, although no one in government will be able to see this because of their silo-based lines of sight. Unfortunately, the policy is being pushed most fervently in a service where some of the poorest in society will be worst affected.

Universal Credit

In 2011, Secretary of State for Work and Pensions Iain Duncan Smith introduced a Welfare Reform Bill which aimed to overhaul the benefits system. At its centrepiece was a single Universal Credit that would replace several benefits: working tax credit, child tax credit, housing benefit, council tax benefit, income support, income based jobseeker's allowance and income-related employment and support allowance (House of Commons 2011). The stated aim of the Universal Credit was to smooth transition into work by reducing the support a person received at a consistent rate as their earnings increased:

> 'The complexity of the current benefit system makes it slow to react to changes in people's circumstances, meaning that many are afraid to try work, but also makes it expensive and difficult to administer. This fuels error by administrators and claimants alike and reduces benefit take-up as people do not understand their entitlements. It is central to the purpose of Universal Credit that it is a simpler system than the one it replaces.' (DWP 2011)

These principles appear sensible and the authors would wish to see the reforms succeed in their aim of 'making work pay' (Daily Telegraph 2011b). However, the Universal Credit was chosen as the first major project to be made to fit in with the Cabinet Office's cross-government 'digital by default' policy. With welfare spending being one of the largest and most politically sensitive areas of public expenditure, it was inevitable that a government focused on deficit reduction would seek to cut the cost of benefits administration. According to the Department for Work and Pensions' (DWP) White Paper:

> 'In delivering Universal Credit, the DWP will adopt the 'digital first' principle and meet the growing demand for flexible and comprehensive online services. For people making Universal Credit claims, notifying changes or checking their payments and responsibilities, the digital channel will be the primary contact route ... The DWP recognises that there will continue to be a minority of people who cannot use online channels. For these people we will offer alternative access routes, predominantly by phone but also face to face for those who really need it. We expect these alternative access routes to be reserved for the minority who can't use, or be helped to use, online services and therefore kept to a minimum.'(DWP 2010)

Here is the first of two fundamental weaknesses in the plan for delivering Universal Credit: it is dependent on the successful construction of a large-scale computer system. How likely is this to be achieved? In their aptly-titled book, 'Dangerous Enthusiasms', Gauld and Goldfinch show that up to 30 per cent of major IT projects fail completely while a further 60 per cent go far over budget and/or fail to meet specifications (Gauld and Goldfinch 2006). In the specific case of the UK, the House of Commons Public Accounts Committee (PAC) noted:

> 'The lack of IT skills in government and over-reliance on contracting out is a fundamental problem which has been described as a "recipe for rip-offs". IT procurement has too often resulted in late, over-budget IT systems that are not fit for purpose ... The UK has been described as "a world leader in ineffective IT schemes for government". There have been a number of high cost IT initiatives which have run late, under-performed or failed over the last 20 years including: the Child Support Agency's IT system, the IT system that would have underpinned the National ID Card scheme, the Defence Information Infrastructure Programme, the implementation of the Single Payments Scheme by the Rural Payments Agency, and the National Offender Management System (C-Nomis).' (House of Commons PAC 2011).

By September 2011, the Daily Telegraph was warning that Universal Credit was top of Chancellor George Osborne's at-risk list of potentially failing projects, with the PAC chairman calling the plan 'a train crash waiting to happen' (Daily Telegraph 2011a). Unsurprisingly, concern centred on the IT build-out. The Child Poverty Action Group's chief executive Alison Garnham told Public Servant magazine: 'The DWP is not famous for its success with new IT projects, and I really worry, because everything goes into one system', adding, 'If it fails it will have really serious consequences' (Public Servant, 2011).

The DWP has remained confident that it can deliver its side of the Universal Credit system on time (Daily Telegraph 2011a), pointing out that the IT contractors are carrying out the project using 'agile' means. Agile – a technique which supposedly allows for the building of computer systems in an iterative, 'live' way – has been described elsewhere as 'merely to do

the wrong thing faster' (Seddon 2011). It is not just the UK that is putting all its eggs in the IT basket. As can be seen from the agenda for the 13th International Conference on Information and Communication Technology in Social Security, the ingrained assumption is that social security services require 'complex administrative tools', 'data exchange between agencies', and hence 'data security' and 'inter-operability' (ISSA website 2012) – all testimony to the scale arguments of the IT companies.

Internationally, there is an almost unquestioned belief that social security administration can only be carried out through a factory-style operation, with pervasive use of IT to deliver both scale and the complexity of benefits that follow. But this is the second and perhaps even more fundamental weakness of current plans. Even if the Universal Credit IT system is built, the service will fail. This is because taxation, credits and benefits are high-variety services that computers are poorly equipped to handle. Using computers to deliver the service depends on codifying in rules the eligibility and entitlements of claimants of every type and condition. Since the rules can never completely specify all the possible variations, the predictable consequence will be poor-quality, hard-to-get services for the most vulnerable in society. The provision will not match the need. The knowledge and awareness of local context that are critical for dealing with claimants' individual circumstances will be lost as local authority housing benefits offices are slimmed down or closed and claimants are instead pushed through the DWP system (DWP 2010).

A revealing glimpse into the distressing way the government intends to treat benefit claimants in order to fit their scale designs was provided in September 2012 (The Guardian 2012b). In preparation for the Universal Credit, the government had set the DWP a target that 80% of new claims for unemployment benefit must be made online by September 2013. The DWP, in its determination to force all benefit claimants to use online services, therefore began deliberately putting claimants on hold when they rang DWP call centres, in order for the claimant to listen to injunctions to go on line. If the claimant were to hold on in order to eventually speak to a call centre agent, then the agent would make a further attempt to convince the claimant to switch to an online service.

This aggressive treatment of claimants is driven by an obsession with cost. Paradoxically, such behaviour toward claimants will inevitably drive costs up. The assumption is that web-based transactions will be cheaper. However, this is to confuse transaction costs (which will indeed be lower online) with the true costs of service, which are determined by the total number of transactions it takes for citizens to get a service. The failure of the web-based service to resolve individuals' problems will multiply those transactions, generating massive amounts of failure demand as vulnerable citizens keep returning until their needs are met. It is most unfortunate that what we have called scale thinking has been allowed to impede an initiative with the potential to improve the life-chances of the poorest in society. The political imperative to dictate not just the 'what' but also the 'how' of policy has meant that the Universal Credit is destined to fail. How best to deliver a policy ought to be treated as an empirical question, not one of managerial ideology. However, decisions about the delivery of the reform have been taken based not on careful learning and experimentation but instead based on a set of management nostrums about how best to cut costs.

What do we discover when we study housing benefits as a system?

Failure demand is a systems concept. Absent any understanding of the conceptual foundations behind the idea, the opportunities it provides for radical improvement will be

missed. As a 'product', the Universal Credit is obviously different from anything that has gone before. However, we can learn much about how to deliver it from experience in housing benefits. Having analysed their housing benefits service as a system, many local authorities have delivered radical improvements; similar results could be achieved if the designers of the Universal Credit followed the same procedure.

In the early years of the New Labour government, Gordon Brown, then Chancellor of the Exchequer, famously asserted there would be no investment in public sector services without reform (The Economist 2007). The DWP persuaded him to invest £200 million in promulgating a new design for the delivery of housing benefits. It was one of the earliest attempts to impose a separate front/back office design in public services. The front office provides the means of access or front door, while claims are processed at the back, the two being connected by electronic work flow in which documents received at the front are scanned and transmitted electronically to the back. Included in the design were targets for both front (how quickly people should be seen, phones picked up and documents scanned and sent to the back office) and back offices (how quickly correspondence should be responded to and claims processed, and how many work activities carried out).

Targets:

- how quickly people are seen
- how quickly phone calls taken

Targets:

- speed of response to correspondence
- time to calculate claims

Front Office

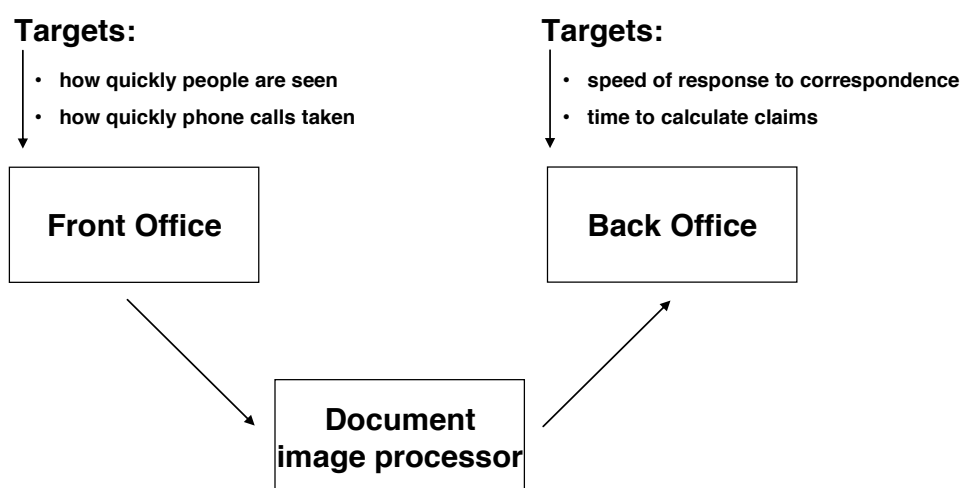**Back Office**

Document image processor

Figure 1: The DWP-sponsored design for housing benefits

When housing benefits services are studied as systems, the flaws in this design become quickly apparent. Analysing demand reveals that, unsurprisingly, very few people present with all the documentation needed to establish eligibility and entitlement to hand. But to conform to front office targets and service standards, whatever is brought in is processed and sent to the back office. This leads to mounting backlogs as back office workers struggle to find and bring together the various electronic 'work objects' belonging to each case into a single 'work object' that can be used to do the 'value' work (establish eligibility and entitlement). Since floating work objects are frequently lost, people find themselves being asked to bring in documents again that they had already produced before. It is not uncommon to find that 20 per cent of objects in the electronic system are duplicates. Meanwhile, to meet their activity targets, back office workers send out letters and forms requesting more information rather than attempting to resolve the claim. Because there is no continuity or case

ownership, each new piece of information requires the case to be started all over again, so the worker has to ask: does this complete the picture, generating still more information requests. The back office becomes a repository of electronic 'work objects' and the job of the workers is to process a number of work objects each day. Much of this work is 're-work', adding no value. As the work accumulates in the back office, it creates increasing volumes of failure demand in the front office too.

One consequence is that claimants return to the front office repeatedly to get their problems solved, increasing tension and lowering staff morale (hence the posters in many offices threatening dire consequences in the case of aggressive behaviour towards staff). Another is that, to tackle massive back office backlogs generated by compliance with the DWP design, local authorities have not only been encouraged to spend tens of millions of pounds employing private sector 'backlog-busters' to search for lost information (i.e., to solve the wrong problem); they have also been exhorted by DWP, the Audit Commission and Gershon to share and/or outsource back office housing benefits services (Audit Commission 2008). Reflecting the government's unshakeable belief in economies of scale, sharing back offices makes matters worse, adding to already high costs and poor service. Indeed, having a back office itself is a design mistake (Seddon 2011).

The first step in understanding the system is to look at demand from the claimant's point of view. There are two types of demand: 'value demand' – 'demand we want', i.e. that the service is set up to provide for – and 'failure demand'. In the case of housing benefits there are only two value demands: 'Can I make a claim?' and 'I have had a change in my circumstances'.

Failure demand includes progress chasing ('What has happened to my claim?', 'I don't understand how to fill in your form'), having to bring in duplicate documents, making repeat visits – all of which create demands on the system, and therefore more work, caused by a failure of the service to work from the claimant's point of view. We find that failure demand typically accounts for as much as 70 per cent of the total demand into housing benefits services – all of which are in full compliance with the DWP's directives.

When they analyse demand, managers are invariably shocked to discover that despite meeting all their targets, service quality from the claimant's point of view is appalling. The end-to-end time for claimants to receive a service typically averages 50 days and can rise to six months. A queue at a benefits front office will often contain people who are turning up for the fifth, sixth and up to tenth time. Making employees work harder is no solution – it is the design that is the cause of poor performance, and at the heart of the design is management by targets. There is a paradox here. Working to targets actually makes performance worse. But meeting targets is the way to win approval from government inspectors.

The second step in the Vanguard Method is to design the service against demand. The idea is simple but profound: the service should be designed to make it easy for the customer to 'pull value' (get what they want) from the system. Knowing the type and frequency of demand – why citizens make demands on the housing benefits service from their point of view – puts people in a position where they can design a service that works.

As we have noted, there are only two value demands in a housing benefits service: 'I want to make a claim', and, 'my circumstances have changed'. We shall use the former – 'I want to make a claim' – to illustrate the approach. The claimant demand dictates the value work – the

things that need to be done to satisfy that demand. In this case the value work consists of:

- Obtaining 'clean' information (a complete picture of the claimant's circumstances)

- Making a decision

- Notifying the claimant

- Paying the claimant (if entitled).

Knowing the type and frequency of demands for claims, it makes practical sense to equip the people at the first point of contact with the necessary expertise to respond to high-frequency predictable demands – demands which the organisation knows it can expect. That is to say, it makes no sense to train everybody in everything; but it does make sense to train people to handle the bulk of the predictable work. When more unusual demands hit the system, people working at the point of contact 'pull' expertise from others equipped to handle such cases. This design principle keeps the ownership with the person providing the service and, consequently, speeds up their rate of learning – as they meet more 'less frequent' demands they develop the skills and knowledge required to deal with them[1].

In practice, many benefits claims are tied up with claimants' council tax obligations. It follows that the expertise required to deal with the council tax implications should be designed into the roles of those who work at the point of transaction with the claimant (which also happens to be something that the current design for the Universal Credit will be unable to do, having been designed to be administered by different government agencies to council tax benefit (Adam and Browne 2012)).

Finally, to understand how well a service is working it is necessary to measure achievement of purpose from the claimant's point of view. This means precise calculation of end-to-end times, from when claimants first present to the point where they receive the correct money or are told they do not qualify. Taking the Vanguard approach, an authority will typically process all benefits within three to six days, a result so far beyond the government benchmark of 28 days that no one would dream of setting it as a target.

Unfortunately, councils following this approach have found themselves in trouble with government bodies. The inspectors inspect against the specifications, and in these designs, however successful in performance and cost terms, they cannot 'tick' their boxes (target-monitoring, plans for sharing services, etc). All too often inspection coerces managers to do the things inspectors have been sent to monitor, not the things that meet customer demand; to fail inspection did not look good in the then Labour government's 'targets and terror' regime

---

[1] Systems designs challenge the current norm of 'dumbing-down' service (hire cheap people, give them scripts and computer-based diagnostics). Traditionally workers are trained against procedures, subjects and rules. This is why, despite the time it takes, conventional training often leaves people both poorly equipped to answer calls and understandably nervous. Every one of us has had experience of talking to call-centre workers who are unable to solve our problem. Instead, people who deliver services need the expertise required to identify and deal with the complexity and variety of customer demands. If they have been trained against demand, staff know exactly how to get help for calls for which they are not trained. Training against demand engages the workers in solving the problems that benefits claimants present with, ensuring that there is continuous improvement. However, the consequence of such 'smartening up' is a fall in costs – the organisation learns to do the 'value work' and only that, driving waste out of the system.

(Hood 2006).

It is salutary to contrast the performance of these authorities with that of the departments which have followed the various scale-based reform edicts of government. As noted above, HMRC's operations represent the most egregious example of the failure of command and control, scale-based thinking. HMRC's amalgamation by then-Chancellor Gordon Brown was in itself an example of official obsession with scale and centralisation. In a classic case of service industrialisation based on the belief that taxation service can be mass-produced like a manufactured product, HMRC has gone 'lean' and gone wrong (Seddon 2012). Their work has been standardised and specialised. The management's focus is on activity, not achievement of purpose. In a misguided attempt to create a performance culture, workers are set to solving management's wrong problem – 'why didn't we meet our targets yesterday?'

While HMRC managers assure parliamentary select committees that lower transaction costs will bring improvement, the evidence is of mounting failure demand (House of Commons Treasury Select Committee 2011, BBC 2010, Financial Times 2011). Accountants have built websites to complain about the number of transactions it takes to get a service. Even when callers to HMRC's call centres can get through they are left uncertain about the advice they have received (Daily Mail 2011). Advice UK, the umbrella for welfare advice organisations, has established that it is costing member organisations across the country at least £500 million to mop up failure demand downstream from HMRC and the equally dysfunctional DWP (Advice UK 2008).

HMRC exemplifies the fallacy of scale thinking: the belief that improved efficiency can be achieved through lower transaction costs, staff specialisation and work standardisation, using an IT 'solution' as the means. It fails because it mistakenly assumes that the fact that some channels are inherently cheaper than others is more important than the design of the system. Unfortunately, the lower cost of transactions is more than wiped out by rising failure demand, which means that the total number of transactions it takes customers to get a service goes up. Based on the central idea that costs are associated with activity, service managers bear down on exactly the wrong lever and drive their costs up.

Despite the evidence from 'flagship' programmes such as HMRC, however, scale assumptions remain all pervasive. A telling example is a recent National Audit Office (NAO) report detailing the failure of UK scale-based shared service ventures, which then manages to draw all the wrong conclusions (NAO 2012). Replete with compelling evidence of the cost of failure, the report concludes (where have we heard this before?) that the problem is implementation and that the answer is to do it better. In particular shared services have failed, according to the report, because providers have mistakenly 'customised' services for clients instead of forcing them to standardise in the cause of both economy and, even more alarmingly, the requirements of the IT platforms – a remarkable example of reversing the proper position of cart and horse. Ignoring its own evidence, the report concludes that Cabinet Office pressure on departments to share services, combined with stronger governance and project management, will ensure the failures are avoided in the future (The Guardian 2012a). This is a good example of 'doing the wrong thing righter', as Russell Ackoff put it, which actually makes things 'wronger' (Stern 2009).

Service is not the same as manufacturing

As we have noted above, many of the attempts to industrialise service come from a

misguided attempt to reproduce the scale economies of Adam Smith's pin factory and Henry Ford's car plants. Manufacturing can survive command and control management (albeit at a cost) because products are standard; there are economies of compliance. But service differs in crucial respects from manufacturing. Aside from the obvious lack of physical manufacture, a service is co-produced by the customer and the service agent. Variety of demand is inherently greater. Instead of thinking of the system as one that pulls physical parts together to manufacture at the rate of customer demand, as in manufacturing plants modelled on the Toyota Production System (Ohno 1988), the system must be conceived as one that brings (largely) intangible expertise together in response to the variety of customer demands. A different purpose requires different methods, since there are different problems to solve. Solving these problems shows how to design services from which customers can 'pull value' (i.e. get what they want).

When applied to service organisations, the traditional command and control scale design responds to the variety of customer demands by establishing procedures, standard forms, functions, levels, specialised 'factories' and the like to deal with it. The consequence is enormous amounts of waste. To eliminate the waste it is necessary to dismantle the functional structures and 'put variety in the line'.

Conventionally-schooled managers believe that this will take time. This is partly because of how they think about change. It is also because of what it implies for their past identity and effort. In the name of 'service' many of our organisations have been built as mass-production factories subjected to detailed programmes of activity directed by managers. Often this represents a significant investment in human and financial resources. To undo or redirect this effort represents a considerable psychological as well as intellectual challenge.

Maximising the ability to handle variety is central to improving service and reducing costs. The systems approach employs the ingenuity of workers in managing and improving the system. It is intelligent use of intelligent people; it is adaptability designed in, enabling the organisation to respond effectively to customer demands. Workers are connected with customers in self-organising relationships. In managing flow the work itself is the information, and this in turn comprises the information required to direct operations in the work.

People are good at handling variety. Computers are not. As managers develop the systems approach, they learn to use computers for the things they are good at and *a contrario* avoid using computers for things that people are good at. The consequences are fewer computer systems and more control. The value of IT lies in supporting those who deliver service; in scale designs, IT systems dominate and hinder the delivery of good service. Computers have become the cement for command and control management, reflecting the unquestioned assumption that managers should set targets and then create control systems – incentives, performance appraisals, budget reporting and computers – to keep track of them. On the contrary, to make service organisations work better, it is necessary to take these things out.

Evidence-based policy

The redesigned benefits service examples demonstrate better service to claimants, at much lower cost, as well as better relationships with people who need support. There are further positive unintended consequences: benefits staff have learned to see their job as helping people to solve their problems (e.g. with housing services, other related benefits such as free

school meals, health services), not just administering benefits – a perspective it would be impossible to take in a fragmented scale design (Middleton 2010). In addition there is clear evidence that fraud detection is far more effective in these local designs than in the nationally-organised framework.

Returning to the 'core paradigm' embedded in the proposals for the Universal Credit, we contend that it contains three major flaws:

- It treats all demand as 'work to be done' ignoring, *inter alia*, the importance of understanding failure demand

- It assumes that individuals should be held accountable for their work. In fact, as W Edwards Deming taught, 95 per cent of individual performance variation can be shown to be attributable to the way the work is designed and managed - the system - and is thus the responsibility of management (Deming, in Scholtes 1998 p296). Holding individuals 'accountable' results in sweatshop conditions (DWP call centres are described as 'living hell' by staff who carried out a two-day strike (The Guardian 2011)), while HMRC is disfigured by a long-running dispute on performance measurement (The Guardian 2008) and disastrous staff relations

- It fails to absorb the variety of service demand. By ignoring the nature of demand and measuring and managing activity, as above, it ensures that costs will rise and services worsen. In both HMRC and DWP, work has been standardised on the assumption that standardisation leads to lower costs; in fact, along with specialisation, it is a primary cause of failure demand and thus raises costs.

Under government proposals, full implementation of Universal Credit will take seven years and a massive investment in IT. This is a hopeless formula for absorbing variety and will – as it always does – generate enormous amounts of failure demand, citizen dissatisfaction and cost. By contrast, if what constitutes the Universal Credit could be defined today, housing benefits offices redesigned along the lines described above could provide it quickly and efficiently in a matter of weeks. In fact, to ensure that the Credit is fit for purpose, these offices should be used to develop the rules, taking risk out of the solution.

The essence of the better design is i) to provide the necessary expertise at the first point of contact to satisfy all of the predictable 'value' demand (a claim or a change of circumstances); ii) to allow agents to 'pull' expertise for the less predictable demand, using measures that relate to the purpose (right money to right people as quickly as possible); and iii) to switch management's focus from managing activity to managing the whole system's achievement of purpose. Using these joined-up principles, housing benefit offices have subsequently learned another lesson. People's needs and problems come in a variety of interlocking forms and guises; solving them all at first point of contact offers huge potential (if usually invisible) cost savings since it reduces knock-on demand on other services. In contrast with factory designs such as HMRC and DWP, morale in services designed on these lines is invariably high because people are intrinsically motivated, illustrating Herzberg's dictum that the best way to get people to do a good job is to give them a good job to do (Herzberg 1987).

However, despite the growing evidence of both failure of scale (e.g. in Western Australia[2], the UK Research Councils[3] or the Department of Transport[4] shared service disasters) and success (Vanguard Method designs have been successful and documented in the UK, New Zealand (Middleton 2010), Netherlands and Sweden), policy-makers stick to plausible but wrong ideas. Unfortunately – and this is the deep sadness of our political system – politicians do not 'do' evidence. Politicians do narrative; they worry more about how to handle fierce television interviewers than whether their initiatives will work. As Alex Stevens (Public Servant 2011) has shown, Whitehall[5] is a policy-making machine that prefers certainty to accuracy, action to contradiction, and only accepts 'evidence' that fits the story already being told. Policy is based on systematic distortion of evidence. Hence the continued promotion of 'scale' ideas despite overwhelming evidence of failure. The costs associated with failure, massive at any time, are even more glaring in a period of austerity. But never mind the evidence of previous IT-based scale failures, 'digital by default' fits the government narrative. Politicians argue for the need to hold local public services to account, but in this context accountability is reduced to little more than compliance with centrally promulgated ideas. As the Finnish educator Pasi Sahlberg puts it, 'Accountability is something that is left when responsibility has been subtracted' (The Atlantic, 2011).

What we have learned from working with public services in various countries is that the structure of government (whether devolved or centralised) is not the critical factor. What does make a difference, however, is the amount and nature of control exercised through specifications – with which services must comply – and inspection, the means for ensuring compliance. Where any agency mandates matters of method and measures, the inevitable problems ensue. The ability to mandate method and measures has to be removed from all agencies and made the responsibility of managers who deliver the service. This change to the locus of control is an essential prerequisite for innovation.

One element of control is the requirement for 'improvement' to be planned, resulting in massive investment in bureaucracies of interpretation, planning, project management, administration, report-writing and self-publicity, all of which are a huge stumbling block to innovation. As all the pioneers described here have learned through experience, change is an emergent property, so that having a plan for it is a contradiction in terms. The priority for management is to study their systems, to get knowledge; to understand how their current work designs sub-optimise performance; to discover the importance of understanding real customer demand; and to learn how to design a system to serve it. En route they discover the counterintuitive truth that economy comes from flow, not scale.

---

[2] Western Australia's Department of Treasury and Finance Shared Service Centre promised savings of $56 million, but incurred costs of $401 million (Western Australia Economic Regulation Authority 2011)

[3] A National Audit Office report said that the UK Research Councils project was due to be completed by December 2009 at a cost of £79 million. But, in reality, it was not completed until March 2011, at a cost of £130 million. (NAO 2011)

[4] The Department for Transport's Shared Services, initially forecast to save £57m, is now estimated to cost the taxpayer £170m, a failure in management that the House of Commons Public Accounts Committee described as a display of 'stupendous incompetence'. The most recent evidence of the higher cost was documented in a House of Commons Transport Select Committee report (HoC Transport Select Committee 2010)

[5] Whitehall is recognised as the centre of Her Majesty's Government. The street is lined with government departments and ministries; the name "Whitehall" is thus also frequently used as a metonym for overall British governmental administration, as well as being a geographic name for the surrounding area.

To conclude, this paper calls for more evidence-based policy. In public services all over the world, there is a ferment of activity in which small groups of dedicated people are enthusiastically demonstrating that better economics follow from delivering services that are truly local. Many examples of success have been published (Middleton 2010, Pell 2012) and more are to come in books and case studies (The Systems Thinking Review website). This is real evidence, but the concern is that it is so much against the grain of opinion and current political narrative that it seems impossible for many minds to comprehend. Many governments, with their ingrained command and control, top-down perspectives, cannot or are unable to accept this evidence. Whilst this paper has focused on the flaws of the particular social security reforms planned for the UK, what is crucial is that governments internationally recognise that the centre must be able to support innovation where it happens locally, rather than crushing it through specifying methods for delivery. Our hope would be for governments to declare a single policy initiative: henceforth responsibility, the prerequisite for innovation in the delivery of services, will be placed firmly where it belongs, with local service leaders.

References

Adam S and Browne J 2012 'Reforming Council Tax Benefit' Institute for Fiscal Studies Commentary C123, London

Advice UK 2008 'It's the System, Stupid! Radically Rethinking Advice' AdviceUK: London

The Atlantic 2011 'What Americans Keep Ignoring About Finland's School Success' Anu Partanen Dec 29

Audit Commission 2008 'Back to front: Efficiency of back office functions in local government'

BBC 2010 'HMRC 'at melting point'' Radio 4 Today Programme Thursday, 9 September (See http://news.bbc.co.uk/today/hi/today/newsid_8982000/8982769.stm accessed 12/3/12)

Cabinet Office 1999 'Modernising Government' The Stationery Office (See http://www.archive.official-documents.co.uk/document/cm43/4310/4310-sm.htm accessed 12/3/12)

Cabinet Office 2005 'Transformational Government: Enabled by technology' The Stationery Office

Cabinet Office 2010 'Digital by default proposed for government services' 22 Nov (Downloadable from http://www.cabinetoffice.gov.uk/news/digital-default-proposed-government-services accessed 12/03/12)

Chase, R.B. 1978, 'Where does the customer fit in a service operation?' Harvard Business Review, Vol. 56 No. 4, pp. 137-42

Computerweekly.com 2012 'Cosy cloud coterie snuggles into top nob gov IT jobs', January 27 (See http://www.computerweekly.com/blogs/public-sector/2012/01/cosy-cloud-coterie-snuggles-in.html accessed 12/03/12)

Daily Mail 2011 Tax chaos will last until 2013, admits HMRC boss as he tells MPs service is 'not acceptable' 17 March (See http://www.dailymail.co.uk/news/article-1367049/Tax-chaos-2013-admits-HMRC-boss-tells-MPs-service-acceptable.html#ixzz1uYMQN7yl accessed 12/3/12)

Daily Telegraph 2011a 'George Osborne is warned of disaster over welfare reforms' 25 Sep 2011

Daily Telegraph 2011b 'Incentive to work has been cut, says Iain Duncan Smith' 5 December 2011

Department for Work and Pensions 2010 'Universal Credit White Paper' The Stationery Office

Department for Work and Pensions 2011 'Universal Credit Policy Briefing Notes – An Introduction' (see http://www.dwp.gov.uk/docs/ucpbn-intro.pdf accessed 12/3/12)

The Economist 2007 'The day of the spider' April 26th (See http://www.economist.com/node/9073969 accessed 12/3/12)

Financial Times 2011 'HMRC sees big rise in complaints upheld' February 27 (See

http://www.ft.com/cms/s/2/dbff6da2-42a4-11e0-8b34-00144feabdc0.html#axzz1uYNbo4oN
accessed 12/3/12)

Ford, H 2003 'Today and Tomorrow' Productivity Press. Cambridge, MA. First published 1926

Gauld, R and Goldfinch, S 2006 'Dangerous Enthusiams. E-government, Computer Failure and Information System Development' Otago University Press: New Zealand

The Guardian 2008 'In the name of efficiency' Monday 9 June

The Guardian 2011 'Working for government call centres makes you sick' Friday 21 January

The Guardian 2012a 'Shared services: centres are over-budget and ERP overly complex' Wednesday 7 March

The Guardian 2012b 'Unemployed deliberately held in call centre queues to promote website' Sunday 16 September

Hood C 2006 'Gaming in Targetworld: The Targets Approach to Managing British Public Services' Public Administration Review July-August

House of Commons 2011 'Welfare Reform Bill' The Stationery Office

House of Commons Transport Select Committee report 2010 'The performance of the Department for Transport: Government response to the Committee's Fourth Report of Session 2009–10' Third Special Report of Session 2010–11, The Stationery Office

Herzberg F 1987 'One More Time: how do you motivate employees?' Harvard Business Review: Sept-Oct 1987 reprint with commentary of 1968 original (vol 46: no 1: pp53-62)

House of Commons Public Administration Select Committee 2011 'Government and IT-"a recipe for rip-offs": time for a new approach' Twelfth Report of Session 2010–12 The Stationery Office

House of Commons Treasury Committee 2011 'Administration and effectiveness of HM Revenue and Customs Select Committee' Sixteenth Report of Session 2010–12 The Stationery Office

HM Treasury 1998 'Comprehensive Spending Review - Modern Public Services for Britain:

Investing in Reform' The Stationery Office

HM Treasury 2004 'The Gershon Review: Releasing resources to the front line: Independent Review of Public Sector Efficiency'

ISSA website 2012 Advert for 13th International Conference on Information and Communication Technology in Social Security (See http://www.issa.int/News-Events/Events/13th-International-Conference-on-Information-and-Communication-Technology-in-Social-Security accessed 12/3/12)

Johnston R 2005 'Service Operations Management: Return to Roots' International Journal of Operations and Production Management, Vol 25 No 12 First published in 1999

Krafcik, J.F. 1988 'Triumph of the Lean Production System' Sloan Management Review,

Vol. 30 No. 1, pp. 41-52

Lane Fox, M 2010 'Letter to Francis Maude' October (downloadable from http://www.cabinetoffice.gov.uk/sites/default/files/resources/Martha%20Lane%20Fox%27s%20letter%20to%20Francis%20Maude%2014th%20Oct%202010.pdf accessed 12/3/12)

Levitt, T. 1972 'Production-Line Approach to Service' Harvard Business Review, September-October

Lockyer, K.G. 1962 'Factory Management', Pitman, London

Middleton, P (ed.) 2010 'Delivering Public Services that Work (Volume 1): Systems Thinking in the Public Sector Case Studies' Triarchy Press: Axminster

National Audit Office 2011 'Shared services in the Research Councils' HC 1459 21 October

National Audit Office 2012 'Efficiency and reform in government corporate functions through shared service centres' HC 1790 7 March

Ohno, T 1988 'Toyota Production System' Productivity Press: Portland, Oregon. Translated from Japanese original, first published 1978

Pell, C 2012 'The Vanguard Method in the Public Sector: Case Studies Delivering Public Services that Work Vol.2' Triarchy Press: Axminster

Porter, M 1980 'Competitive Strategy' Simon and Schuster: New York

Public Servant 2011 'Policy is based on systematic distortion of the evidence' 12 April 2011 (See http://www. publicservice.co.uk/feature_story.asp?id=16180 accessed 12/3/12)

Scholtes, PR 1998 'The leader's handbook: making things happen, getting things done' McGraw-Hill, London

Seddon, J 2003 'Freedom from Command and Control', Vanguard Press.

Seddon, J 2008 'Systems Thinking in the Public Sector: the failure of the reform regime and a manifesto for a better way', Triarchy Press.

Seddon, J 2010 'Why do we believe in economies of scale?' (http://www.systemsthinking.co.uk/6-economies.asp accessed 12/3/12)

Seddon, J 2011 'Universal Credit: a brilliant idea guaranteed to fail' (See http://www.universalcredit.co.uk/universal-credit-a-brilliant-idea-guaranteed-to-fail/#comments accessed 12/3/12)

Seddon, J 2012 'Academia can tax one's patience' Public Money & Management Volume 32, Issue 2, pages 127-128

Smith, A 1998 'The Wealth of Nations' Oxford University Press: Oxford.

Stern S 2009 'Fond farewell to a brilliant thinker' Financial Times 09/11/09 (http://www.ft.com/cms/s/0/0168c7de-cd7e-11de-8162-00144feabdc0.html#axzz22Hyp8bBK accessed 01.08.12)

The Systems Thinking Review website www.thesystemsthinkingreview.co.uk

Taylor, FW 1998 'The Principles of Scientific Management' Dover Publications: New York. First published in 1911

Western Australia Economic Regulation Authority 2011 'Inquiry into the Benefits and Costs Associated with the Provision of Shared Corporate Services in the Public Sector- Final Report' Perth WA

Zokaei Z, Elias S, O'Donovan B, Samuel D, Evans B and Goodfellow J 2010 'Lean and Systems Thinking in the Public Sector: Report for the Wales Audit Office' Lean Enterprise Research Centre: Cardiff University