

The Toby Lowe Letters

Foreword and Summary

This paper constitutes an exchange of e mail correspondence between David Burnby, a leading UK Results Based Accountability™ (RBA) practitioner and trainers and Doctor Toby Lowe, Chief Executive Officer of Helix Arts and a visiting Fellow at Newcastle University Business School. It also features input from Mike Pinnock, an experienced practitioner in children and young people's services who has written and co-authored a number of publications on outcome-based work and related subjects.

The exchange of correspondence arose out of work Dr Lowe has undertaken into the use of outcomes as a framework to measure the impact of social policy interventions. His position can be summarised thus:

- The use of outcomes as a concept to measure the effectiveness of social policy interventions is inherently flawed and creates unwelcomed paradoxes
- Evidence suggests that focusing on outcomes distorts both the priorities and practice of organisations
- The use of targets based on outcomes inevitably leaves to data manipulation and game play disadvantaging those at most need
- The complex web of relationships surrounding service user outcomes makes it virtually impossible to trace outcomes back to activity and it is therefore unreasonable to hold service providers accountable for outcomes
- Data is not the same as impact: data that is simple to measure and collect will not in itself capture the complexity of lived experience
- Using outcome measurement as the basis for "Payment by Results" approaches creates particularly insidious distortions

David acknowledges the validity of most of these arguments. The fundamental difference in opinion surrounds the extent to which adopting the OBA approach will inevitably lead to the consequences summarised above. David's responses are in essence a defence of the OBA approach. His position can be summarised thus:

- The conclusions Dr Lowe draws arise out of his misunderstandings around the principles and practice of the OBA approach, in particular the subtleties around accountability and how data is used to improve performance
- Many conclusions that Dr Lowe draws are based around a single piece of academic research from Australia which is itself inherently flawed
- OBA is not a crude logic model that relies wholly on data, nor does it advocate simplistic target setting around outcome measures as the basis for evaluating performance
- Mark Friedman, the originator of RBA, has developed the "Next Generation Contracting" model as the basis for commissioning outcome focused services which is not solely reliant on data, but advocates amongst other things the establishment of performance partnerships between purchaser and provider
- Views around the potential pitfalls of "Payment by Results" approaches are shared.

The Toby Lowe Letters

Correspondence between **David Burnby** and **Dr Toby Lowe: 2012-13**

This paper is made up of correspondence exchange and papers between David Burnby and Dr Toby Lowe. For ease of reading, responses are colour coded for the avoidance of any doubt about the relative authorships.

Dr Toby Lowe is Chief Executive of Helix Arts, a Participatory Arts organisation which works with the most disadvantaged and marginalised people in society. Prior to Helix Arts, he worked for Arts Council England, and the Social Exclusion Unit in the Office of the Deputy Prime Minister. He has a background in political philosophy, particularly the way in which “community” is used a concept in politics. Toby is also a Visiting Fellow at Newcastle University Business School, undertaking research into the use of outcomes as a framework to measure the impact of social policy interventions

Prior to becoming a freelance Facilitator, Trainer and Consultant in 2002, **David Burnby** had over twenty years experience working in the voluntary sector, initially in the field of community enterprise and worker co-operative development. Following five years as a CVS Chief Officer, he worked for eight years for the national leadership development organisation Common Purpose, originally as Programme Director and latterly as Regional Director.

Since 2004, David has specialised in supporting organisations and partnerships using the Outcome Based Accountability™ framework developed by Mark Friedman in the United States. David has trained and worked with Mark and is now regarded by him as a leading OBA/RBA practitioner in the UK. David has worked intensively supporting outcome focussed approaches with several Local Strategic Partnerships, NHS Bodies, Health and Wellbeing Boards, Children and Young People Partnerships and Local Authorities across the UK.

This correspondence came about as a result of David Burnby attending one of Dr Lowe’s lectures “*Kittens are Evil: Heresies in public policy*” in October 2012. Dr Lowe argues that the use of outcomes as a concept to measure the effectiveness of social policy interventions is inherently flawed and creates unwelcome paradoxes. Instead of improving the lives of service users, his cites evidence suggesting that focusing on outcomes distorts both the priorities and practice of organisations, leading to poorer results for those most in need. ***In his lecture, he cited Outcome/Results Based Accountability™ (OBA/RBA) as being culpable in this regard.***

The correspondence essentially challenges this alleged culpability. David argues that OBA/RBA is being misrepresented in Dr Lowe’s work whilst Dr Lowe robustly defends his thesis. Surprisingly, there is a significant amount of common ground between the two protagonists: both are opposed to concepts such as payment by results and ‘command and control’ target setting management thinking. Both have the interests of the end user at heart. But whereas Dr Lowe sees OBA/RBA as a process that compounds the worst effects of outcome centred performance management, David argues quite the contrary. He goes into some detail to describe the finer points of OBA/RBA which readers who are unfamiliar with the process might find helpful. Dr Lowe introduces concepts such as complexity theory and argues that introducing accountability into the outcomes mix will inevitably lead to unintended consequences, gaming and goal displacement.

It should be noted that RBA (Results Based Accountability™) is known as “Outcome Based Accountability™” (OBA) in England and Northern Ireland only. The two expressions refer to the same framework, but for the avoidance of confusion, David refers to RBA™ consistently throughout his comments.

The conversation opened in November 2012 with Toby responding to an invitation from David for further contact following his “Kittens are Evil” lecture in Newcastle in October 2012. This is the first reference to a piece of research published in Australia by Keevers et al to which various references are made later in the text.

From: Toby Lowe
Sent: 06 November 2012 16:36
To: david@davidburnby.co.uk
Subject: Reference

Hi David

Apologies for not following up our conversation sooner. After I finished my presentation I went straight into organising and delivering an 8 day festival – so have only just emerged from the other side of that. Ah, the joys of a portfolio life!

Here’s the reference I referred to in my paper that specifically focuses on Results Based Accountability (which, I believe is what they call OBA in Australia). It’s the first study I’ve seen of its kind – doing detailed ethnographic research onto the before/after effects of running an RBA process on the practice of frontline workers. It really is very good.
<http://oss.sagepub.com/content/33/1/97.short>

It presents evidence about the way in which RBA processes undermine relationship building between support workers and young people.

So -whilst you suggested that OBA practice escapes the problems of program-logic models and target regimes (and other aspects of using outcomes-based evaluation as performance management mechanisms), the evidence from this paper suggests that the kinds of impacts it creates on frontline practice fit the same pattern of distorting impact as the Perrin, van Thiel, and Bevan studies I cited relating to other types of outcomes-based performance management frameworks.

I am keen to investigate further the distinctions you raised between “population outcomes” and (a phrase I didn’t catch!). Could you explain what you meant by this?

Many thanks, and I’m looking forward to further conversations,

Kind regards

Toby Lowe

From: David Burnby [mailto:david@davidburnby.co.uk]
Sent: 06 November 2012 17:49
To: 'Toby Lowe'
Subject: RE: Reference

Hi Toby, and thanks for getting back to me. To clear up potential confusions around the labels, the process Mark Friedman developed is known as “Results Based Accountability” (RBA) everywhere in the world except for England and Northern Ireland where the expression “Outcomes Based Accountability” (OBA) was adopted. Both names have been trademarked in order to protect the unique characteristics of the framework which distinguish it from other so-called outcome frameworks.

The issue around accountability is fundamental. RBA/OBA distinguishes two distinct types of accountability:

Whole Population Accountability is the accountability communities have to the well-being of whole populations (e.g. “All people in Newcastle”). We can only bring about outcomes (i.e. conditions of well-being) for whole populations through a range of actions and interventions from different agencies and organisations. So we cannot reasonably hold the NHS (for example) wholly accountable for the health and wellbeing of all people in Newcastle, or the police wholly accountable for Newcastle being a safe community. We need to work in partnership with a range of organisations and sectors to develop strategies (i.e. collections of actions) that will contribute towards whole population outcomes. We should hold such partnerships accountable for the effectiveness of these strategies by monitoring progress against outcomes.

Performance Accountability is the accountability of service providers for the well-being of their clients. So if part of our Health & Wellbeing Strategy (for example) is to provide a smoking cessation service, we would hold the managers of that service accountable for the impact the service has on its client population. In OBA/RBA, we determine the data to performance manage the service based on three questions: How much did we do?; How well did we do it?; Is anyone better off? The answer to the third question is the client outcome which in turn contributes towards the whole population outcome.

The reason the distinction is so important is that it differentiates between Ends and Means. Whole population accountability is about Ends (the desired conditions of well-being we want for our communities) and performance accountability is about the means of getting there. If the two are confused (and they often are), then we can get trapped into an ‘all talk and no action’ loop of endless meetings and little progress.

In both Whole Population and Performance accountability, we use data to inform progress. But because the data is playing different roles, it’s important it has different labels. So data used to inform progress against a strategy at the whole population level is generally known as “indicators”, and data used to inform the effectiveness of services on their client populations is generally known as “performance measures”. The terminology varies, but the important thing is that there is a discipline of language use so we can be sure what the data is telling us.

I suggest you re-visit Mark Friedman’s book “Trying Hard is Not Good Enough”¹ for a fuller description on accountability and the RBA/OBA process.

¹ “Trying Hard Is Not Good Enough” by Mark Friedman published by FPSI publishing 2005 ISBN 1-4392-3786-7

I'll certainly read up on the Australian research so I can comment. You also mentioned some research carried out in Newcastle which I'd be interested in seeing. I think you also suggested that you had recently published a paper in a journal which I didn't catch the name of which again I'd be interested in reading.

Kind regards,

David Burnby

From: David Burnby [mailto:david@davidburnby.co.uk]

Sent: 25 February 2013 11:54

To: Toby Lowe

Subject: How outcomes Saved My Life

Hi Toby

I hope this finds you well. A colleague of mine, Mike Pinnock, brought to my attention your article in the Guardian recently, and it gave rise to an interesting discussion. Mike is a fellow proponent of the Outcome Based Accountability™ approach and we realised that there was a danger of us being 'guilty by association' of some of the distortions of outcome thinking that your work is focussed on. You'll recall at your presentation in Newcastle last year I said that I agreed with 95% of your thinking in terms of how management approaches such as the adoption of crude target setting, or payment by results contracts are in the main divisive and counter-productive. It was, and remains the inclusion of this type of deficient management practice in the same breath as 'outcomes' that concerns me. Outcomes don't kill people; managers do. Mike recently penned the [attached article](#) which was published in the New Zealand "Social Work Now" magazine. He wrote it without knowledge of your work, but it does touch on some common ground argued from a different perspective. As ever, I'd be interested in continuing this dialogue with you, not least to ensure that the focus of your work attacks the misuse of outcome measurement rather than its proven benefits.

David

From: Toby Lowe

Sent: 25 February 2013 16:59

To: 'David Burnby'

Subject: RE: How outcomes Saved My Life

Hi David

I enjoyed [Mike's article](#) – and I think it gets to the heart of a key problem concerning outcomes – their use as performance management tools. And I wholeheartedly agree that when 'outcomes' are used as targets, they're just as destructive and ridiculous as any other form of target-based management.

I also agree that a discussion of desired end states - between partners and those they are seeking to serve - is part of bringing together effective partnerships with common purpose. And that trust is key to thinking about alternative approaches to performance management.

And all this generates a number of questions in my mind...

- Am I right in thinking that an 'evolved' OBA/RBA approach would recommend that the process of devising and agreeing 'outcomes' is the most important element of the process? and that whatever this process comes up with, it shouldn't be used to set targets on which to monitor the performance of people/teams/organisations/programmes?
- If that's right, what are the performance accountability measures? How are they related to the desired outcomes? How do complex outcomes (which are unattributable to any one particular intervention) relate to performance measures for a particular organisation?
- And how do performance accountability measures escape the problems of management by targets? (The evidence I've seen about RBA's use as a performance accountability tool in Australia indicates that it generated accountability mechanisms that were detrimental to good frontline practice – the Keevers et al article)
- How does this process respond to some of the limitations that thinking about impact in terms of outcomes creates, e.g.:
- How does thinking about desired outcomes respond to complexity? In most complex systems, there is no 'golden thread' (Mike's words) which you can trace from desired outcome back to activity. It's a 'golden web' of complex relationships and emergent properties. Doesn't encouraging people to think about a 'golden thread' (or any other version of program logic models) encourage people to think that problems are actually simpler than they are? Don't we need to embrace complexity to respond effectively to difficult social problems?
- How can problems of measurement be avoided? Manipulation of data is only the most obvious of the measurement problem. A far more crucial point seems to be that any data that is collected will not be the same as understanding impact. (Because any data that is simple and efficient to collect by definition won't be capturing the complexity of lived experience). So – what is the role of data in respect of both population and performance accountability measurements? Isn't one of the problems of thinking about 'outcomes' in this way: that they then *demand* to be measured? (What's the point of saying we want to create an outcome that 'children are to be safe' if we don't measure whether that outcome has been achieved or not?) And as soon as we have to measure, we're faced with the problem that we can either measure how people have genuinely been effected by an intervention (which is v costly, time consuming and likely to produce incomparable, context-dependent results) or we can produce simple data which enables us to compare change over time and between places and different interventions. (How can the 'Uncertainty Principle' be resolved?)

Looking forward to hearing your thoughts.

Cheers

Toby

Ps – to answer your question about focus – I'm doing a lot on 'Payment by Results' as the exemplar of the problems of Outcome Based Performance Management. Did you see the piece I had on the Guardian Professional network about it?: <http://www.guardian.co.uk/local-government-network/2013/feb/01/payment-results-staff-fictions>

At this point, Mike Pinnock, the writer of the article [“How Outcomes Saved My Life”](#) referred to earlier joins in the correspondence to respond to the questions Toby raises in his previous e mail about the “Golden Thread”.

Mike Pinnock has been involved in partnership working for children and young people for the past 38 years both as a practitioner and as a senior manager. In the early Nineties, he was one of the first people in the UK to explore and apply the ideas of outcome-based services for children and their families. Since then he has worked extensively with Mark Friedman of the Fiscal Studies Policies Institute on applying the Outcome-Based Accountability (OBA) framework. He has contributed to a number of national and international projects on outcome-based work. He currently works as an independent researcher/trainer and as a consultant researcher at the Centre for Child and Family Research, Loughborough University. He has written and co-authored a number of publications on outcome-based work and related subjects, the most recent being for the Munro Review of Child Protection.

From: Michael Pinnock
Sent: 01 March 2013 17:51
To: Toby Lowe
Cc: David Burnby
Subject: Golden Thread or Golden Shred?

Hello Toby

Dave sent me a copy of the exchange he's been having with you regarding OBA. I won't steal Dave's thunder - but I will respond to the specific point you make about 'the golden thread'.

Typically, I use this analogy when I'm working with staff and service users on agreeing desired outcomes as opposed to attributing actions to actual outcomes (so upstream rather than downstream). That's the way I've treated it in the article. Basically I'm suggesting that if you've chosen what you believe to be an 'evidence-informed' way of working (and by that I mean one that service users [past, present and potential] believe will help them to live the sort of lives they want to lead), then you should be able to see how your efforts help realise the intended outcomes.

In the final ten years of working in local government I managed a lot of 'enabling' service as well as direct services - so things like staff training, asset management, ICT, business support, complaints and so on. I covered children and families as well as adult social care services - and also a public sector housing. I found that this idea of finding the golden thread that connected your work to better outcomes for service users/tenants worked well on two levels. First it helped people to nurture and protect the "the things that needed to go well to bring better outcomes" ('The gemba' as I believe it's referred to in Deming circles). Second, it helped them to see how the bit that they needed to get right helped 'make a difference'. I'll give you an example.

I used to have a team member called Sheila. Sheila liked to knit, watch Hull City and drink beer - sometimes she'd do all three things at once. She was a greater "sorter" of messy problems. We gave her the job of sorting the foster carer payments system out. Payments were often late, wrong or missed altogether. Understandably, this used to piss them off. Sheila persevered until she'd sorted it out. In doing so she also made a very tangible contribution to improving outcomes for Looked After children. Our relationship with carers improved dramatically when the payments system was sorted out. It meant they had more time to give to their family because they didn't have to waste hours on the phone trying to sort out our incompetence. It also meant that they were more likely to recommend fostering to a friend/neighbour/colleague/relative. This made it easier to recruit and retain foster carers. Ultimately it was one of the things that allowed

us to give children and young people better choice in placements. Sheila wasn't the only one that helped improve the support we gave to foster carers - but she made a significant contribution.

Over the past decade, I've seen this golden thread idea misappropriated by the corporate performance wonks who use it in a very different way. As you suggest, they use it to try and 'hard-wire' a link between specific actions and actual results - a link which in reality is often nothing more than wishful thinking. I've also seen the term used to create spurious links between plans and strategies to demonstrate 'synergy and alignment' (the fetishes of design-school planners). Again, in reality what's holding these plans together is nothing more than a 4mm cold steel staple.

When I'm talking about actual results, I have been known to ask the question 'Golden Thread or Golden Shred?' as a way of exploring the question of attribution, externalities etc. Drawing on complexity theory, we know that outcomes emerge through a complex web of interactions that are affected by and affecting the environment that they are operating within. I've used the Tipping Point ideas to encourage people to think about how they might work to create the conditions for change (the Golden Shred idea fits well with Gladwell's idea of 'stickiness').

To be honest, I've tended to shy away from talking about complexity theory when I'm working with practitioners - people tend to glaze over when they hear the term - so I'm always looking at ways of introducing the way of thinking without actually referring to outcomes as attractors with CASs and so on. Golden Thread/Golden Shred has been my way into this - as has the stuff that I've done on the limitations on the Machine Metaphor.

So I guess that's a long-winded way of saying that I agree with you!

What I've tried to say in my article are two things. First, outcome-orientated approaches are not just about measurement. Second, when we measure things, we often corrupt them (as we learnt this week with the report on Sapphire). As I've said in the article, this is usually nothing to do with what's being measured, but why and how it's measured.

Incidentally, on the subject of 'evidence-informed' government, Peter Smith anticipated all of the likely negative effects of over-zealous outcome-based performance systems (in health care) with unerring accuracy - in the early Nineties.

Mike [Pinnock]

From: Toby Lowe
Sent: 09 March 2013 13:51
To: 'Michael Pinnock'
Cc: 'David Burnby'
Subject: RE: Golden Thread or Golden Shred?

Hi Mike

Thanks for your thoughtful response to my question. And apologies for the delayed response – I was super busy last week preparing for the ‘kittens are evil’ conference. (Which went really well – was a great buzz about the place)

One of the other speakers summed at the conference – Andy Brogan from Vanguard - summed up the issue of measurement really well, I thought. He said that if you measure for learning and improvement, then measurement is good. But if you measure for accountability, then the measure replaces and subsumes your original purpose. And you can either measure for learning OR for accountability, but not both.

At the conference, we launched a campaign against Payment by Results – which is the ultimate in measurement for accountability, I think. If you'd like to support the campaign, and help fight the pernicious nonsense, you can sign the 'say no to PbR' declaration here: www.saynotopbr.net

Cheers

Toby

In this paper, completed on the 24th July 2013, David responds to Toby's e mail of 25th February:

Hi David

I enjoyed Mike's article – and I think it gets to the heart of a key problem concerning outcomes – their use as performance management tools. And I wholeheartedly agree that when 'outcomes' are used as targets, they're just as destructive and ridiculous as any other form of target-based management.

I also agree that a discussion of desired end states - between partners and those they are seeking to serve - is part of bringing together effective partnerships with common purpose. And that trust is key to thinking about alternative approaches to performance management.

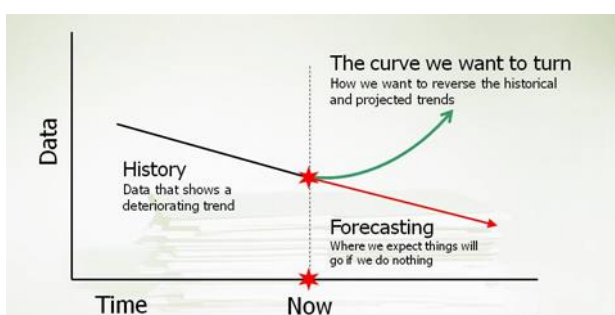
And all this generates a number of questions in my mind...

- Am I right in thinking that an 'evolved' OBA/RBA approach would recommend that the process of devising and agreeing 'outcomes' is the most important element of the process? and that whatever this process comes up with, it shouldn't be used to set targets on which to monitor the performance of people/teams/organisations/programmes?

OK. before I can answer this, I need to again stress the need to separate out accountability. The start of the RBA process is for lead authorities or community leaders to work with partners and stakeholders to decide on the conditions of well-being (outcomes) they desire for whole populations. So, for example, "All people in Newcastle (a whole population) enjoy good health and well-being". This is surprisingly unique in strategic planning. Most processes start with a needs analysis or list of problems, a deficit based approach which attempts to design services to fill in gaps and fix problems. It will almost inevitably see service users as passive recipients of services rather than potential agents of change in their own lives. The RBA approach believes in starting with where communities want to be (as

defined by a series of outcomes), and then designing strategies to move towards that outcome. I believe that it is preferable to move towards an outcome than away from a problem.

We then need to define these outcomes with a handful of **indicators**; e.g., obesity rate, prevalence of smoking, all-age, all-cause mortality etc. It's important here to take account of the use of the term "Indicator" to describe data that informs progress towards a whole population outcome. We use a different term ("**Performance Measure**") to describe data used to inform the performance of Service Providers. The distinction between the two and the discipline in use of language to describe them is critical in maintaining the distinction between population and performance accountability.



Using the partners' understanding and knowledge of their own communities, and guided by best practice, the final role is to come up with a strategy, (or a collection of actions with a reasoned chance of success of making a contribution to the desired outcome).

The RBA approach advocates measuring progress at the whole population level using data presented as a line graph in three parts: the historical trend of the given indicator, a prediction of the future trend if no action is taken (the red line) and a trend line (green) to show the desired direction of travel. Progress is indicated by distance travelled, or the rate of the curve turned. As a general rule, Friedman does not advocate setting targets for the rate of distance travelled in the belief that this creates a negative management culture and almost inevitably leads to game play, unethical behaviour and data manipulation². The accountability at this level lies with the partnership responsible for the strategy.

With the strategy agreed, accountability for the service providers for each agreed action lies with the managers of those services (or those responsible for the actions) who are accountable for the impact of their service on their customer or service-user populations. We refer to this as performance accountability. The distinction is critical. It's the difference between means and ends. Partnerships define strategy (or the ends); managers deliver services (the means). Confusion between the two results in endless meetings with all talk and little action. The RBA approach again advocates using data to measure progress on the basis of distance travelled, not point-to-point improvement (i.e. targets).

- If that's right, what are the performance accountability measures? How are they related to the desired outcomes? How do complex outcomes (which are unattributable to any one particular intervention) relate to performance measures for a particular organisation?

² See 'Trying Hard is Not Good Enough' Page 87-89 for Friedman's view on the use of targets

RBA defines three types of performance measures answering three basic questions on how the services impacts on customer or service user populations: How much did we do? How well did we do it? and Is anyone any better off? The last measures can be described as customer or service-user outcomes (as opposed to whole population outcomes). Traditional public sector commissioning focussed almost entirely on “how much” measures in the mistaken belief that providing lots of service automatically means people will be better off. RBA focusses particularly on the “better off” data whilst acknowledging the importance of quality (how well?) measures. Again, incentivising improved performance through arbitrary target setting is actively discouraged. Progress is measured by the rate of curves turned and distance travelled.

The relationship between service-user performance measures and whole population indicators is a contributory one. We can hold managers accountable for the impact their interventions make on their client population (as measured by the ‘better-off’ performance measures), but in any system of complexity, no manager can sensibly be held accountable for the well-being of a whole population. Cause and effect relationships are almost impossible to understand.

This is why maintaining clarity between whole population and performance accountability is critical. We cannot tell what part any individual service or intervention has on a whole population outcome, so we need to stop pretending we can! But we can measure the effectiveness of a complete strategy as a collection of actions. If we implement a series of interventions at a given point in time, and we have the evidence at the performance accountability level that service interventions are making an impact on their customer populations, and after that point in time we saw a turn in the curve at the whole population level, then we are demonstrating the effectiveness of our strategy. It suggests that we’re doing the right things.

Friedman described this to me as akin to the parent who helps her/his child with their school homework. If the child succeeds in their education (by reaching a measurable level of attainment) then it would be fair for the parent to declare that they contributed towards that outcome, but nonsense to suggest their input was the sole factor of success. It would also be a fruitless exercise to attempt to calculate precisely what proportion of the child’s success was attributable to the homework support factor. What we could conclude would be that a series of quality actions (good teaching, parental support, absence of bullying, healthy diet and lifestyle etc.) applied through the child’s education collectively contributed towards the successful outcome. So it’s about doing the right things (as determined by the planning process at the whole population level) and then doing those things right (as determined through performance managing actions or interventions)

- And how do performance accountability measures escape the problems of management by targets? (The evidence I’ve seen about RBA’s use as a performance accountability tool in Australia indicates that it generated accountability mechanisms that were detrimental to good frontline practice – the Keevers et al article)

I don’t share your enthusiasm for the Keevers article, not least because of its factual inaccuracies with regard to RBA, the description for which bears no measure of familiarity to

me. (I also think it's questionable practice to draw conclusions about the experience of RBA across Australia based on a single piece of academic research focussing on one relatively small voluntary organisation (i.e. Southern Youth)). Given how much store you place on this article, I thought it merited a closer look:

In her introduction³ Keevers places RBA in the mould of an auditing tool for government agencies and throughout her text presents the framework as a purely economic model, distant from community values with the 'policing' of performance as part of a contractual agreement as the prime driver. RBA is described as *"a bundle of material-discursive practices"* which I think I've decoded to mean in this context imposing external pressures on the organisational practices of communities. Given that the RBA process is designed to start with a blank piece of paper, I'm unsure how this conclusion is drawn, but maybe I've misunderstood the term. Keevers lumps together⁴ Friedman's work with that of Hatry et al (Measuring Program Outcomes: A Practical Approach 1996), a behemoth of a manual, with the most of the contents being wholly anathematic to RBA, and Réal Lavergne's paper "Results-Based Management and Accountability for Enhanced Aid Effectiveness 2002" which is an academic study focussing on managing results and accountability and again bears little resemblance to Friedman's work.

And here lies the rub, and the grave mistake you echoed in your early work: that is a failure to recognise that RBA is a trademarked and not generic name. It refers to a very specific process and a collection of quite unique principles as presented in the book "Trying Hard is Not Good Enough" (and various academic papers published by the Fiscal Policy Studies Unit). So Keevers's criticism of RBA is wholly undermined by her comments in the opening preamble of her paper by the expression *"RBA in its various forms (Friedman, 2005; Hatry et al., 1996; Laverge (sic) 2002)..."* which demonstrates her complete misunderstanding of the terminology. RBA has only one form and it has little to do with the work of Harry Hatry or Réal Lavergne. RBA was developed as a means of using data to improve conditions of wellbeing for whole populations and service users. It was not designed as a stick to beat people up with for failing to deliver on arbitrary targets. If government agencies or indeed misguided practitioners choose to use such practices and label them "RBA" then that is the fault of the perpetrator, not the framework itself.

Keevers accurately describes a whole population "Turning the Curve" process⁵ in her article, but then leaps to the conclusion that *"... RBA proposes evaluating the 'results' of practices, without having to focus on the organizing practices themselves. RBA planning thus assumes it is possible to go straight to the end-result and by-pass the practices employed in delivering the services and achieving the 'results'. In looking at outcome as output, RBA attempts to measure the 'effect' of practices"*. This statement demonstrates a complete lack of understanding of the RBA methodology and confusion between Whole Population and Performance Accountability. The process Keevers describes is that of designing strategy at the whole population level; it's about designing a coherent collection of actions with a reasoned chance of contributing towards whole population outcomes. Measuring the

³ Made to Measure: Taming Practices with Results-based Accountability: Lynne Keevers, Lesley Treleaven, Christopher Sykes and Michael Darcy (January 2012) from Organization Studies

⁴ ibid (Page 99 Lines 7-15)

⁵ ibid (Page 107)

'effect of practice' is a different RBA technique which operates along a parallel track to the process Keevers describes. This is about managing and improving performance of actions or services; a wholly different process to designing and monitoring progress towards outcomes for whole populations. Keevers demonstrates little understanding of the distinction between the two and makes a critique of RBA having only witnessed half the process!

There are many other misunderstandings and inaccuracies in Keevers's paper with regard to RBA in practice that questions the credibility of the authors and the validity of its conclusions. On Page 110 for example, Keevers quotes Mark Friedman from an interview with Sandra Handley in "Local" Newsletter (Autumn 2008) where Friedman attempts to define the crucial distinction between accountability for services and whole population accountability. From a brief quotation, taken out of context, Keevers concludes that measuring the effectiveness of performance using RBA is only relevant to "service provision". From this she implies that RBA excludes any form of practice that "cannot be constituted as a service". This is simply not the case from which a modicum of research would identify. For example, in a paper published by the Results Leadership Group in 2009⁶, H. Daniels Duncan (Senior Vice President for the United Way of Tucson and Southern Arizona) presents the case for how Asset Based Community Development (ABCD) allies itself perfectly with RBA. This is an important marriage of ideologies of the type that Keevers fails to recognise. ABCD favours co-production; models of action which are fundamentally different to the conventional model of service provision by statutory providers. ABCD sees communities as assets that are part of the solution; not as passive recipients of "service delivery". Rather than RBA excluding *"...from mattering all the practices that cannot be constituted as 'a service'"* Duncan sees no such contradiction between RBA and ABCD, describing the two as "complementary approaches that can provide the necessary framework to help local United Ways complete this transformation [of communities] and achieve real long-term results"

As I've indicated before, RBA thinking doesn't kill people, managers do. Blaming RBA for the adoption of a target setting culture is akin to blaming the hammer when the thumb is hit. RBA uses data not as a stick to beat people up with, but as a means of improving conditions of wellbeing. The "Turning the Curve" principle of measuring distance travelled is central to RBA thinking and target setting is an anathema to this principle. Poor managers will always look for simple solutions to complex problems and many will use targets in the mistaken belief that they incentivise employees to perform better. All the evidence I've seen suggests this is not the case and service users almost inevitably suffer in order for service providers to make their targets. It is quite wrong to ally the culture of target setting with RBA.

- How does this process respond to some of the limitations that thinking about impact in terms of outcomes creates, e.g.: How does thinking about desired outcomes respond to complexity? In most complex systems, there is no 'golden thread' (Mike's words) which you can trace from desired outcome back to activity. It's a 'golden web' of complex relationships and emergent properties. Doesn't encouraging people to think about a 'golden thread' (or any other version of program logic models)

⁶ Redefining the Role of United Ways with Results Based Accountability and Asset Based Community Development

encourage people to think that problems are actually simpler than they are? Don't we need to embrace complexity to respond effectively to difficult social problems?

Absolutely. Incidentally, RBA is not a logic model. Friedman is very critical of logic models because life is rarely logical. The notion that one course of action will lead logically to a given outcome doesn't take into account the complexity of factors that impact on this type of work. RBA simplifies the planning process and as a result makes it more accessible to a broader range of partners and service users: but it doesn't attempt to suggest that the problems themselves are simple; neither does it attempt to disguise the complexity of the systems it means to influence.

A key feature of RBA planning is that it takes into account the "Story behind the Baseline" before attempting to determine appropriate actions. This is not the same as a 'logical framework' that attempts to predict the cause and effect relationship between every action. The Story Behind the Baseline is the opportunity to engage effectively with a whole range of stakeholders, many of whom will have been alienated from conventional planning exercises. The process recognises the knowledge, experience and skills that lie outside the 'professional' input and in doing so, bring some practical meaning to community engagement. "The Story" is perhaps a misnomer because in reality, it's a whole range of stories from different perspectives. And of course, research and established good practice are also part of the Story.

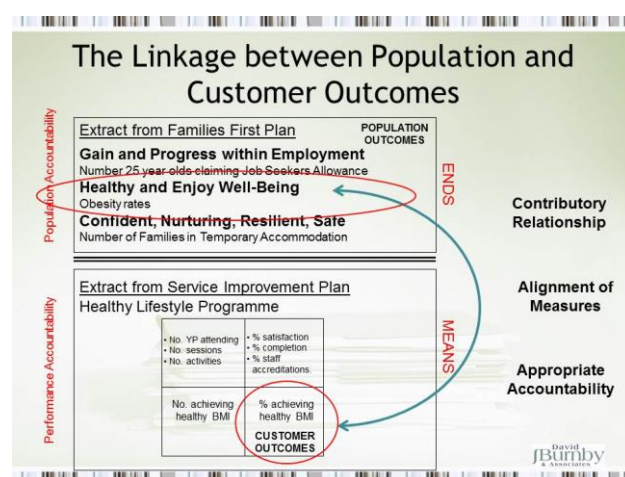
- How can problems of measurement be avoided? Manipulation of data is only the most obvious of the measurement problem. A far more crucial point seems to be that any data that is collected will not be the same as understanding impact. (Because any data that is simple and efficient to collect by definition won't be capturing the complexity of lived experience). So – what is the role of data in respect of both population and performance accountability measurements? Isn't one of the problems of thinking about 'outcomes' in this way: that they then demand to be measured? (What's the point of saying we want to create an outcome that 'children are to be safe' if we don't measure whether that outcome has been achieved or not?) And as soon as we have to measure, we're faced with the problem that we can either measure how people have genuinely been effected by an intervention (which is v costly, time consuming and likely to produce incomparable, context-dependent results) or we can produce simple data which enables us to compare change over time and between places and different interventions. (How can the 'Uncertainty Principle' be resolved?)

My experience of community projects had been first of all trying to fit the needs and aspirations of communities into the criteria of the funder (so that the emerging strategy would be funding rather than needs led); with funding secured, spending six months appointing and inducting staff, two years trying to deliver against pre-determined, usually arbitrary targets; six months winding down and then, in the last breath of the project, hiring an evaluator to see if the project made any difference. Small wonder that the deprivation gap persists. In this model, data served only to provide the illusion of effective management and force priorities which would not necessarily have much relationship to community needs and definitely would not bend to reflect the rapidly changing environment.

At the whole population level, use of data is necessary to define priorities. An outcome of “Children are healthy” is vague and rhetorical. Its purpose is to set the line of travel and ensure that everyone is moving in the same direction towards the same place. For the journey towards the outcome to be meaningful, we need data to measure our distance travelled and to continually assess if we’re moving in the right direction. So we may define healthy children in terms of under 5 mortality rates, body weight (obesity levels at key stages); teenage conceptions etc. We set our aim to ‘turn the curve’ on what we consider to be the most compelling indicators that reflect the biggest challenges in communities. So from day one, we’re gathering data on a handful of indicators (typically no more than three to five for any given outcome) and using that data to measure our progress along the road towards the outcome.

Incidentally, we don’t talk about “delivering” or “achieving” outcomes because we know that every single child in a given community will not be healthy. But we can contribute towards the outcome by a collection of actions with a reasoned chance of success. Evidence through data that we’re turning the curve shows our progress on that destination from the start of the process, not the end. Using data in this way also maintains a focus on the things that matter most. Rather than proposing a huge list of indicators and setting up a massive data collection factory to collect and report on it all, the RBA approach is to focus on a handful of proxy indicators which lead the ‘data herd’ (i.e. indicators that say something of central importance to an outcome which, if moved, will cause others to follow. E.g. If we choose smoking as a health indicator, if we turn the curve on numbers of smokers, then we’ll also turn the curve on heart disease, chronic obstructive pulmonary disease, hypertension, cancer etc.)

The choice of indicators is critical and there is always a balance to be struck between the proxy power of the indicator and the availability and practicality of collecting timely and accurate data. I don’t accept that indicators that say something fundamental about the quality of peoples’ lives would necessarily fail to be prohibitively difficult to collect. But I concede that collecting data around outcomes can be costly and challenging. But if we are focussing on a handful of critical data sets, it has to be a good investment. Otherwise, what is the alternative? Blindly going ahead with a strategy without having a clue as to whether or not anyone is any better off?



Much of the same arguments apply to performance management data, except that because you are in most instances dealing with a discrete and much smaller client or service user population, data is easier to collect and can more accurately reflect the complexity of an individual’s life experience. For example: If we are commissioning a healthier living programme because we want to contribute towards a whole population outcome of say “Health and Wellbeing”, then we need to be clear about what

impact we want the programme to have on the people it works with that will contribute towards that outcome. We would collect a limited amount of data to answer the questions “How much did we do?”; “How well did we do it?”; and, critically, “Is anyone better off”. These measures of being “better off” should align and contribute towards our whole population outcome by addressing at least one of the priority curves to turn (indicators). The graphic illustrates how population and performance accountability fit together in this way.

I understand what you mean about the “uncertainty principle”, but we’re not dealing with quantum mechanics here! If we’re setting up or commissioning projects to make people healthier, there will not be a single ‘magic bullet’ performance measure that will tell us we’ve delivered on service user outcomes. Performance Measures are always about approximation and compromise. But if we run a healthy living programme based on three performance measures, say (for example), Body Mass Indicator, Blood Pressure and Alcohol Consumption, and the data tells us that 85% of our service users experience improvement in these areas, then that’s good enough to say that this is a project worth investing in which will contribute to a whole population outcome of a Healthy Community (and, in the longer term, an indicator such as longevity).

Of course data can be manipulated. Microchips can be cloned but it doesn’t stop us using credit cards. We can reduce the likelihood of data manipulation by creating a culture where we use data to help us improve our services and communities, rather than to punish us for deemed ‘failure’. Part of that is desisting the practice of setting meaningless targets which, in my experience, do little to improve service user outcomes in anything other than the very short term.

Following this response, David became aware of an article “*New development: The paradox of outcomes—the more we measure, the less we understand*”⁷ written by Toby Lowe that had been published in “Public Money and Management”. In this article, Toby cites the case of “Ryan” as “Ryan’s Story”:

Ryan’s Story

Interviews with people who have experienced multiple problems in their lives and have succeeded in generating positive ‘outcomes’ clearly demonstrate the complexity of their journey. An illustration of this can be seen in the case of Ryan, the client of a supported-housing organization. He was a drug user with a complex family background. Ryan is 18, and was involved in criminal behaviour for a significant portion of his teenage years. Ryan is now settled in supported living, and has ceased his criminal activity.

Ryan’s journey to this point was very complex, and bears little relationship to a simple program logic model. He had interventions from substance misuse teams, housing bodies, youth offending teams and employment support for five years. When enabled to tell the story of the interventions which had helped him turn his life around, the key factor he described was the support he was offered by support workers with whom he had developed

⁷ Toby Lowe (2013) New development: The paradox of outcomes—the more we measure, the less we understand, Public Money & Management, 33:3, 213-216, DOI: 10.1080/09540962.2013.785707 (May 2013) <http://dx.doi.org/10.1080/09540962.2013.785707>

a relationship over time: ‘someone who believed in me’. Crucial to this support was the fact that it was not withdrawn when Ryan ‘failed’ — when he ended up back in court. He cited the example of a particular housing worker who accompanied him to court, swore positive statements for him, and arranged for Ryan to attend educational courses which kept him out of prison. He also attributed a significant element of his turnaround to his girlfriend, who helped persuade him to stop offending.

David followed up his previous comments with some further points arising out of this article in a paper sent to Toby on 2nd August 2013:

From: David Burnby [mailto:david@davidburnby.co.uk]
Sent: 02 August 2013 15:36
To: Toby Lowe
Subject: The Paradox of Outcomes

Hi Toby

I’ve just read your article “New development: The paradox of outcomes—the more we measure, the less we understand” as published in ‘Public Money and Management’ (April 2013). It an interesting and compelling article and provides a useful basis for further discussion.

I’ve put some further observations in response to your paper in the attachment. As ever, I look forward to a continuing debate around some of the issues you’ve raised.

David

Comments on an article by Dr Toby Lowe published in ‘Public Money and Management’ (April 2013)

Some of my observations are covered in my last correspondence where I address specific questions you’ve had about the OBA™/RBA™ approach, not least the blurring of lines between outcomes for whole populations and outcomes for individual clients of services which are still evident in this article. But I’m more interested in your comments around impact on individuals and the hypothesis that outcomes for service users “*can only be understood in relation to the complexity of their lives*” and what you present as the paradox “*Information about outcomes can either be simple, comparable and efficient to collect, or it can be a meaningful picture of how outcomes are experienced by people. It cannot be both*”. I think this is something of a false dichotomy.

I agree that there are many influences which impact on the improvement of quality of life in an individual; your case study (“Ryan’s Story”) makes this point well. And it also demonstrates how we can tie ourselves up in knots if we blur the accountability lines.

Creating a positive outcome for Ryan requires a partnership approach from a range of agencies, organisations and individuals. It is doubtful if any one element of this partnership working in isolation could bring about a positive outcome for him. So once we’re clear about the outcome we want for Ryan (which we could articulate perhaps as “Ryan is outside of the criminal justice system, free of drugs and living in stable accommodation”) and how we would measure our distance travelled towards that outcome, we need to design an appropriate strategy for interventions that have a reasoned chance of success. Our strategy would be informed by our understanding of Ryan’s

story from a range of interested parties (including Ryan and his girlfriend) who would be best placed to use this knowledge to decide the most appropriate interventions. For each of these interventions, we can measure performance by asking three questions: How much did we do; how well did we do it; and is anyone better off. Data used to inform the last question helps us understand our progress towards our client outcome.

If I was commissioning a service provider as part of this strategy, I'd want some basic data to be satisfied that the provider was making a difference to Ryan's life. Simply the fact that the service was being provided ('how much did we do') and evidence to suggest it might be being done well (how well did we do it?) is not sufficient to indicate if Ryan is better off. (We could be funding the delivery of a whole load of inappropriate interventions delivered very well). And this is why there is a renewed focus on measuring outcomes; certainly not because it's easy or simplistic; quite the contrary. It's because for years we've delivered and commissioned services without the slightest clue whether or not they were making a scrap of difference to the people whose lives we were trying to improve. I've been an accomplice in this waste of time and resources; locked into performance management regimes of reams of data around 'inputs' and 'outputs' as though improving quality of life for people like Ryan was a simple, mechanistic process rather than a complex web of inter-agency and partnership working. Which is why I question the efficacy of logic models (of which RBA™ isn't one). On that, I guess we can agree.

But I don't agree that measuring outcomes is in itself a barrier to improving Ryan's life. If I'm commissioning an agency to improve the quality of life for Ryan and people like him, I'd want to make it clear that I'd be looking to see him outside of the criminal justice system, in stable housing and weaned off his drug misuse. These are the outcome measures. And they would only be achieved (as in your pen picture) by all the agencies working together with Ryan and the people close to him. We can measure the effectiveness of the agency's contribution by applying these performance measures to their client population but as you rightly say, in the web of complexity that is Ryan's life, it would be impossible to say with any conviction which particular intervention was directly and solely accountable for these outcomes. But what we can say is that we had a sound strategy: the collective actions of partners made a measurable difference to Ryan's life. And we can be reasonably assured that the agencies we commissioned to contribute towards that process are making a difference to their clients' lives and helping to contribute towards our defined outcomes. (In simple terms, we can evidence that we did the right things, and we did them well). And that's the best we can do, and I believe it's good enough. And I'm utterly convinced that it's a whole lot better than simply commissioning agencies to deliver services on the basis of how much they do (or outputs).

You quite rightly point out that because of external forces and drivers, there is much about improving Ryan's life that is outside the scope of any one individual or service provider. I have a different answer to your question *"What happens when managers are asked [to] manage under these conditions?"* I completely agree that target setting will almost always result in game play, manipulation of data and distortion of service delivery. Which is why Friedman cautions against it and why I use every opportunity to dissuade commissioners from resorting to what I feel is a wholly inappropriate management tactic.

My answer is “*the good managers work in partnership with others*”. For example, if we’re commissioning the prison service to make a difference to their service users’ lives, our “better off” measure might be the recidivism rate. We are aiming to make offenders better off by addressing their offending behaviour and keeping them outside of the criminal justice system. But; would it not be reasonable for the prison service to say that there are far too many influences impacting on the offenders’ behaviour outside of the prison service which makes it unreasonable to hold them accountable? Well, if government is only interested in commissioning a prison service to lock hundreds of people up (how much did we do?); and make sure they don’t escape (“how well did we do it”) then they’re off the hook! But if we see the prison service as a means of making people better off by challenging their behaviour, then prisons can only improve their outcome measure (recidivism) by working in partnership with others. This means working with health providers to challenge alcohol and drug misuse; with businesses to create employment opportunities on release; with education providers to address numeracy and literacy issues. And with the offenders’ family amongst others.

This is effective leadership: clarity about the outcome and what success looks like and an ability to keep people on track and moving in the same direction. Deming’s quote in your final paragraph is self-contradictory in that it attributes the very worst of management practices to a focus on outcomes. The understanding that outcomes are central to bringing people together to establish common ground to pursue a common objective is a critical leadership skill. How can you lead if you can’t articulate your outcome, define and measure progress towards it? I agree there is a paucity of leadership in the public sector which is why there is so much experience of managers wandering around in ever decreasing circles, obsessed with archaic and redundant micro-management practices that wouldn’t understand an outcome if it bit them in the backside! This is precisely why I adopted RBA™ as my preferred leadership tool; to help counter this frustrating and fruitless practice.

You will not be surprised that I am saddened by you continuing to cite the Keevers et al article as ‘evidence’ of how the RBA™ process in itself distorts outcomes. My previous correspondence covers this in more detail, safe as to say I view this as a sloppy piece of research by an author who clearly misunderstands RBA™ and thinks it legitimate to blame the tool rather than the operator.

The consequence you quote⁸ raises a very real fear, and one I’m acutely aware of. We will not serve vulnerable people well by setting up management systems that disproportionately divert staff time away from the needs of clients. The Keevers quote states *‘The integration of this new system has significant impacts on the daily practice of workers, reducing the time available to create a sense of belonging and develop young people’s skills that will form the basis for sustaining their life changes’*. This in itself demonstrates how the RBA™ process has been misinterpreted. RBA™ is not a ‘system’ in this sense. It’s a thinking process to help inform the effective use of resources towards improving outcomes. It is a completely flawed analysis to cite RBA™ as the reason why services to clients were compromised when in fact it was clearly the inappropriate interpretation and misuse of the process that was to blame.

⁸ *Both the managers and the youth workers claim that the accountability and monitoring requirements of the program...[are] paradoxically making it more difficult to engage with and build relationships with homeless and at risk young people. The integration of this new system has significant impacts on the daily practice of workers, reducing the time available to create a sense of belonging and develop young people’s skills that will form the basis for sustaining their life changes (Keevers et al., 2012).*

All this said, I welcome your article and the debates it surfaces. Measuring outcomes is not a straightforward process and I would warmly encourage you in your work to consider the most effective ways of steering the ever reducing resources we have towards the ever increasing needs of our most vulnerable communities.

We share a lot of common ground; I just hope I'm convincing you that RBA™ is an ally in your quest, not an enemy!

David Burnby
02.08.13

Toby Lowe responds to David's two submissions with a paper on the 13th August 2013:

I'll start with thoughts on your response to the 'Paradox of Outcomes' paper. I really enjoyed this response because it focusses on the practicalities of providing support to Ryan. We have a whole load of theoretical issues to discuss, so it's great to begin by grounding it in a sense of why this debate matters so much.

I'd like to focus on the problem of using outcomes as a basis for commissioning, which I think are highlighted by the response you make. You suggest that an outcomes-commissioning framework is a good way to ensure that Ryan gets the services he needs. You say...

"Creating a positive outcome for Ryan requires a partnership approach from a range of agencies, organisations and individuals. It is doubtful if any one element of this partnership working in isolation could bring about a positive outcome for him. So once we're clear about the outcome we want for Ryan (which we could articulate perhaps as "Ryan is outside of the criminal justice system, free of drugs and living in stable accommodation") and how we would measure our distance travelled towards that outcome, we need to design an appropriate strategy for interventions that have a reasoned chance of success"

And...

"If I was commissioning a service provider as part of this strategy, I'd want some basic data to be satisfied that the provider was making a difference to Ryan's life. Simply the fact that the service was being provided ('how much did we do') and evidence to suggest it might be being done well (how well did we do it?) is not sufficient to indicate if Ryan is better off"

I take it from this that you're saying that you would commission a support service for Ryan which would use data about whether or not he had offended, whether he was in stable accommodation, and whether he was taking drugs or not as the 'impact' element of performance accountability. I also take it (from Friedman's work) that this impact element is the most significant of the four quadrants of performance accountability.

So, from this, my assumption is that if, after some specified period of time, (let's call it two years, for the sake of argument) Ryan is not behaving in the desired way (he is offending, or his tenure is unstable, or he is still taking drugs) then the service would not be deemed a success, and it would, in all likelihood, not be recommissioned. Is this a fair reading of what you're saying?

If it is, then this illustrates nicely why I think that such an approach is harmful to Ryan. When Ryan told us his story of the services that had made a difference to him – that had enabled him to turn his life around – the point that he made repeatedly was that what made a difference for him was that the same person supported him and believed in him, even after he ‘failed’. He had been receiving a range of support around his housing, his drug use and his offending behaviour, but he kept getting into fights, and therefore kept winding up back in court. He described going round this circle many times, and for a sustained period. However, his support worker, with whom he built up a strong relationship over time, kept believing in him, kept returning to court and arguing on his behalf, and trying things to help him. He described how this relationship, this sustained belief, made a difference to him in the end, because it made him believe in himself.

If the organisation which was supporting him had been commissioned using the outcomes-approach that I believe you’re advocating, then that relationship would likely have been broken. Ryan wasn’t achieving the desired outcomes. Why would a commissioner have stuck with that organisation? Even if we assume that Ryan achieved the desired outcome within the contract period, what if it hadn’t been that time that he escaped his loop? He escaped on the 5th or 6th time around the circle. What if it had been the 10th or the 20th?

No organisation (or combination of organisations across a programme) can be accountable for Ryan’s behaviour. There are too many external factors. In his interview, Ryan also described that one crucial factor in him stopping his offending was that he got a long-term girlfriend. What if that was the factor that meant he was in a position where he was finally able to listen to, and accept the help from, the support worker he trusted? What Ryan needed was continuity of support, from professionals able to use their best judgement about what Ryan needed. What he didn’t need was an organisation having to stop supporting him because they’ve had their funding pulled because he wasn’t achieving the desired ‘outcomes’.

In your other response paper, you say that RBA provides the framework for managers (and presumably commissioners) to learn and improve so that they can get better at what they do:

“Of course data can be manipulated. Microchips can be cloned but it doesn’t stop us using credit cards. We can reduce the likelihood of data manipulation by creating a culture where we use data to help us improve our services and communities, rather than to punish us for deemed ‘failure’.”

If this is what I thought RBA was about, then I would find it significantly less problematic. But it’s not called ‘Results Based Learning’; it’s called ‘Results-Based Accountability’. Above you stated that “as a commissioner... I’d want some basic data to be satisfied that the provider was making a difference to Ryan’s life”.

In this situation, your basic data is telling you that the service isn’t making a difference. Ryan keeps offending. His drug use hasn’t stopped.

You say that you don’t want to use this data as “a stick to beat people with”? But what does this actually mean for the commissioning process? (I think you’re being unrealistic about what happens

when people use ‘outcomes’ data for commissioning purposes). If you’re commissioning on the basis that the data should demonstrate impact, and the data is not demonstrating impact, what happens?

If (as with every outcomes-based commissioning process I’ve ever seen) this impact-data does change future commissioning decisions (and if you’re not using this data to change future commissioning decisions, what is it for?), then providers will come to know that, and will game the system so as to produce the appropriate data for you next time round. This process was identified by Donald Campbell in the 1970s, when he formulated his ‘Law’ of performance management: “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”

Overall, I feel that your two responses don’t acknowledge the cultural impact of using results-data as a mechanism for Performance Accountability. I agree that what should be happening is that people sit down together and agree strategies, and then use data to learn and improve. But as soon as you use results-data for performance accountability, that’s not what happens.

In the end, the culture of an organisation is set by the performance management framework in which it operates. If commissioners (or managers) are holding their subordinates to account using results-based data (if they are choosing which providers to commission, or deciding which staff members have done a good job (or not)) then that data will not be suitable to use for learning, because the people who produce it will have learnt to manipulate it to demonstrate how well they are doing.

(It is important to note that this manipulation isn’t caused by a moral failing of particular people or organisations. Such manipulation is the only rational behaviour in a system in which you are being held to account by means of something which is outside your control, particularly within systems where you are competing against other organisations for resources.)

So – that’s what I think the practical implications are of using RBA as a performance management tool – why I think it harms people like Ryan: it prevents Ryan from getting the services he needs because his behaviour does not match with the desired ‘outcomes’ that organisations are being paid to create. Now, let me address the theoretical problems that I think cause this to be the case:

Performance vs Population Accountability

In our discussions, you have frequently highlighted the difference between population and performance accountability. And you suggest that is this difference enables RBA to overcome the distorting effects of targets....

“With the strategy agreed, accountability for the service providers for each agreed action lies with the managers of those services (or those responsible for the actions) who are accountable for the impact of their service on their customer or service-user populations. We refer to this as performance accountability. The distinction is critical. It’s the difference between means and ends. Partnerships define strategy (or the ends); managers deliver services (the means). Confusion between the two

results in endless meetings with all talk and little action. The RBA approach again advocates using data to measure progress on the basis of distance travelled, not point-to-point improvement (i.e. targets)."

"As I've indicated before, RBA thinking doesn't kill people, managers do. Blaming RBA for the adoption of a target setting culture is akin to blaming the hammer when the thumb is hit. RBA uses data not as a stick to beat people up with, but as a means of improving conditions of wellbeing. The "Turning the Curve" principle of measuring distance travelled is central to RBA thinking and target setting is an anathema to this principle. Poor managers will always look for simple solutions to complex problems and many will use targets"

I don't think the distinction between population and performance accountability does the work you want it to do. Your argument seems to be that results for groups of service users are acceptable as performance management tools in ways that results for broader populations are not. This seems to be based on the idea that whole population outcomes aren't suitable as performance management tools for organisations, because any one organisation (or programme) does not provide services to the whole population.

This is true, but it's only part of the reason why population outcomes can't be used as performance management measures. The biggest part of the problem is that any intervention that is made in a person's life is only a part of a larger, complex system which operates in order to produce a result. This is as true for a service user group as it is for the whole population. This means that organisations are not in control of results. Friedman acknowledges this: *"the more important the performance measure... the less control the program has over it. This is a paradox at the heart of doing performance measurement well"* http://www.raguide.org/1_1.shtml

But his response to this acknowledgement left me (literally) bewildered.... *"Don't accept lack of control as an excuse... If control were the overriding criteria for performance measures then there would be no performance measures at all."* http://www.raguide.org/3_1.shtml

Correct me if I'm wrong, but isn't he effectively saying "we're going to hold people accountable for things that they're not in control of"? If this is true, why bother to distinguish between population and performance accountability in the first place?

The Impact of Performance Measures on Behaviour

Your responses indicate that you think that the use of Results-Based Performance Measures won't create a target culture. This seems strange to me. RBA uses 'Impact'-based Performance Measures, which are acknowledged to be beyond the control of the people/teams/organisations which are being held accountable. It is therefore bound to generate a target culture. If an organisation (or team, or person) are not in control of whether their actions result in meeting a Performance Measure, they will learn to control those things which do result in the creation of data which meets that measure: cherry picking and parking, teaching to the test, reclassifying data or simply making it up. Let me demonstrate why I think this...

Let's follow Friedman's advice about setting Performance Accountability measures for the bottom-right (impact) quadrant:

"Ask "In what ways could clients be better off as a result of getting this service? How we would know if they were better off in measurable terms?" Create pairs of measures (# and %) for each answer (e.g. # and % of clients who get jobs above the minimum wage)."

http://www.raguide.org/RA/3_7.htm

Say we're running an employment support programme for people who have experienced homelessness. Say we got 25 people into work last year, which was 15% of the people we served. If I understand the RBA process correctly, a reasonable performance measurement objective to set ourselves would be to improve both the number and the percentage year on year (because that improvement is necessary to 'Turn the Curve'). So let's say we want 30 people to get a job next year, which represents 20% of the people we work with.

This looks a lot like a target to me. And if we apply this Performance Measure in a thoroughgoing way, it will have exactly the distorting effects of any other kind of target. So, if we're a commissioner, and we're commissioning on the basis that we want a service to deliver this result, what happens if that service doesn't achieve it? Do we decommission? What happens if they don't achieve the desired numbers and percentage the following year? Or the year after that?

If this Performance Measure is used in this way, it will have exactly the same distorting effects as any other numerical target: the people involved in delivering the service and producing the data will game the system so as to produce the relevant data, rather than providing the service that is needed to these vulnerable people.

If commissioners or managers adopt numerical, results-based distance-travelled targets, and then use them to performance manage those they commission or line-manage, they will create the classic target-culture behaviour.

Let's turn to the example that you gave in your response paper:

"Performance Measures are always about approximation and compromise. But if we run a healthy living programme based on three performance measures, say (for example), Body Mass Indicator, Blood Pressure and Alcohol Consumption, and the data tells us that 85% of our service users experience improvement in these areas, then that's good enough to say that this is a project worth investing in which will contribute to a whole population outcome of a Healthy Community (and, in the longer term, an indicator such as longevity)."

If you commission a service based on changes to service users' BMI, Blood Pressure and Alcohol Consumption, what you will get are programmes that are brilliant at delivering improved data for service users on BMI, Blood Pressure and Alcohol Consumption. Service deliverers will cherry pick clients that they know will show improvements in these measures, whilst rejecting those that they know will be more difficult to make show improvements. They will put their resources into activities which reduce BMI, Blood Pressure and Alcohol Consumption, irrespective of whether that is what

clients actually need. They will attempt to reclassify people to demonstrate improvement – what if we moved the boundaries of alcohol measurement? Finally, if all else fails, they will make the figures up.

We know this is what will happen, because it's what the evidence says **always** happens whenever you use this kind of data as a Performance Measure (see for example Perrin or van Theil and Leuww). In the precise example you give, if an organisation had been funded to deliver a service (and the funding wasn't tied to achieving particular results) then those figures would indeed be useful and helpful to learn what types of intervention may improve wellbeing for that community. But as soon as you use those measures for accountability (e.g. you commission on the basis of those measures), then they become subject to Campbell's Law. Again, if the approach was called Results Based Learning, that would be fine. But in the end, it's interested in Accountability, not Learning. You can't say "we should create a learning culture" and then hold people to account for their livelihoods using the data they produce.

Your responses seem to ignore a fundamental point about the way in which Performance Measures work: they direct performance. If you measure the 'success' of people's activity through Performance Measures, those people will adapt their behaviour to meet the measures of success. (If Performance Measures aren't doing this, what are they for?) This is why any numerical Performance Measure is bound to direct worker's activity away from meeting the needs of the people that present to them. Unless those people's needs exactly match those that have been specified in the Performance Measures, then the use of those Performance Measures will result in a worse service for the people that need it. The difference between what is being measured (what is used as a Performance Measure) and the impact that is felt in people's lives is crucial. This is why the (on the face of it) entirely sensible and pragmatic suggestion that people should be performance managed using available, proxy data is (actually) so damaging. Whatever data is available, this will come to define what 'doing a good job' looks like, irrespective of the actual needs of the people who present to the service.

Complexity and Past Performance

The response that I have seen to this point, in your papers, and in Friedman's writing, is that picking performance measures that relate to established baselines means that these are not arbitrary targets, and therefore it is legitimate to use them in this way.

I don't think this argument stands up to scrutiny at all. Outcomes are emergent properties of complex systems. The key characteristic of complex systems is that if you change just one other variable, then the outcome of that system will be different.

"In complex systems, patterns emerge as a result of multiple interactions between agents and only repeat by accident; they are coherent in retrospect but not in advance" Snowden, D. 'Managing for Serendipity or why we should lay off 'best practice' in KM, ARK Knowledge Management Vol. 6, Issue 8, 2003

Or to put it another way, past performance does not give a reliable indicator of future results. As a really practical example, imagine that the employment programme example I used above

established its baseline figure of 25 people into work in 2007. Then the economy crashes. Of what use is that baseline figure now? Substitute 'the economy' for any of the huge range of variables that are external to programme delivery, any of which can change at any time, and you can see that any numerical Performance Measure is essentially arbitrary. It doesn't matter if this was established using a previous baseline, because there is no way of knowing whether external variables have changed in such a way as to make that baseline irrelevant.

Logic Models

I wanted to respond in particular to your point about logic models. You say:

"RBA is not a logic model. Friedman is very critical of logic models because life is rarely logical. The notion that one course of action will lead logically to a given outcome doesn't take into account the complexity of factors that impact on this type of work. RBA simplifies the planning process and as a result makes it more accessible to a broader range of partners and service users: but it doesn't attempt to suggest that the problems themselves are simple; neither does it attempt to disguise the complexity of the systems it means to influence."

Obviously, you know Friedman's work better than me. But this does contradict what Friedman says about Logic Models on the RBA website. Here he indicates support for Logic Models:

"(2) There is a great value in the work being done under the rubric of "logic model," or "theory of change." Logic models and the Results-Based Accountability model presented in this guide are all, of course, "logical." And in fact both approaches are complementary. But there are important distinctions worth noting:

(a) Logic model work is mostly about program performance and generally does not address whole population well-being. It is a performance accountability method or program evaluation method.

(b) It works best for programs which provide a specific service or set of services. (It is often used to design program evaluations.) It is less useful for whole agencies or for parts of agencies that do not provide service (like administrative functions).

(c) Logic model thinking processes, like the United Way's approach, work in the "opposite direction" of results-based decision making and budgeting. One starts with clients, the other ends with clients.

Logic models start with program resources and then describe activities, work product outputs and finally the ways in which clients are better off, i.e. client results.

Performance accountability skips directly to "who are our clients?" then identifies client results, baselines, the story behind the baselines, partners, what works and an action plan.

These can be seen as complementary thinking processes. It would be possible to use a logic model to identify client results and then pick up with the performance accountability process from client results to action.

(d) Many program logic models do not address quality of effort (upper right quadrant) measures like staff ratio, timeliness, cultural competence etc. This must be added.

(e) Most logic models do not distinguish between the quantity (#) and quality (%) of client results. ("Our program helped get 15 people off of drugs." This number is good if the total number served was 20; and bad if the total served was 1,000. The percentage or rate shows this difference.)"

http://www.raguide.org/RA/3_9.htm

It seems to me that Friedman's critique of logic models is technical, not philosophical. You're right to suggest that Friedman does make nods to issues of complexity. However, the problem he faces is that he can't acknowledge fully the role of complexity because linearity is built into the fundamental principles of RBA. We can see this intrinsic linearity in what Friedman says about starting with ends and working backwards to means: *"Results decision-making uses results...as the starting point for making decisions. It is a business-like process that starts with ends and works backwards to means"*
http://www.raguide.org/1_1.shtml

This seems to me to be the philosophical bedrock of RBA – the idea that you can imagine the results you want to achieve, and then “work backwards” – which must mean working backwards along a logical chain of cause and effect – to identify the actions that will get you there.

The problem with this is that it is impossible. You cannot “work backwards” in complex systems. Where results are the emergent properties of complex systems, ‘working backwards’ simply cannot be done. Which leaves Friedman with a problem – either he properly embraces complexity, or he rejects the idea of “working backwards” from ends to means.

I would be intrigued to hear your thoughts about how the basic idea of RBA can be reconciled with complexity.

The Keevers Article

I'm really interested in the digging that you've done about the quality of data in the Keevers article. In particular, I'm interested in the question that you raise about whether or not her primary research relates to organisations which had undergone specifically RBA planning and accountability processes. I'll ask her (I must be able to find an email address for her), and if it turns out that she is referring to Outcome Based Performance Management processes more generally, I'll change my presentation and writings to reflect that.

Summary

Taking all this together, I think my thoughts about RBA are as follows:

- If RBA was simply a process which brought different stakeholders together to discuss strategies for service improvement, and how they were going to use data to learn, then - you're right – I wouldn't have a significant problem with it. I would probably suggest that the linearity seemingly inherent in its worldview meant that it was not the most helpful way to conceptualise the issues at hand, but the way that you apply Friedman's thinking may help escape this.
- As soon as RBA is used for Performance Accountability, then the real problems kick in. If it is used for commissioning, it will distort commissioning processes so that the most vulnerable people's needs go unmet. If it is used to manage individual's or team's performances within organisations, it will distort those teams' or individuals' behaviours. The key to this problem (and why it isn't a technical issue about what is measured, or how managers interpret data) is that it is trying to make people/teams/organisations accountable for results that they're not in control of - because those results are emergent properties of complex systems. And so, because

Performance Measures direct performance, those people/teams/organisations will learn to control what does produce the data which demonstrates effective performance (cherry picking etc).

So – it's not that I think RBA is 'the enemy' (that's definitely Payment by Results!), I'm just not persuaded that RBA escapes any of the significant problems that I have seen with Outcome Based Performance Management in general. And I think that this matters, because I know that this will result in people like Ryan not getting the services they need.

The key to this is the way in which the 'Accountability' element of RBA operates. Your responses seemed to me to say that data on results for service user groups should be the most important element of a commissioning process. If that isn't the case, then there is scope for further dialogue here, I think.

Perhaps our differences on this are to do with the different experiences that have given us the passion to do the work we do. You have clearly seen the distorting and wasteful effects of output-based performance management practice, and have thought – I can help make a difference to this. I have seen the distorting and wasteful effects of outcome-based performance management practice, and I feel the same way.

The day after this response, Toby sent David the following e mail relating to the Keevers et al referencing:

From: Toby Lowe
Sent: 13 August 2013 14:04
To: 'David Burnby'
Subject: FW: Taming Practice with Results Based Accountability

Hi David

Wow – I got an unbelievably fast response from Lynne Keevers. Below is her response to my question about whether her conclusions were drawn from fieldwork exclusively focussed on organisations that use RBA (as derived from Mark Friedman) or from wider outcomes-based accountability practice:

Our fieldwork was with 5 community based organizations including SYFS (the organization discussed in the org studies paper). Some of the funding agencies to whom SYFS was accountable used outcomes based accountability that was not specifically Friedman's but our study was focused on Friedman's version since the government agencies that funded the organizations in our study were mandating Friedman's RBA at the time.

As a result, I will continue to reference her work in the way that I have been doing.

Cheers

Toby

David's final detailed response to Toby was made in a paper submitted on the 12th December 2013 as an attachment to the following e mail:

From: David Burnby [mailto:david@davidburnby.co.uk]
Sent: 10 December 2013 15:29
To: Toby Lowe
Subject: Third response

Dear Toby

Seasonal greetings to you. How quickly three months flies past. It was August when you sent your last response to me. Being a quiet period for me, I quickly put some responses together but then got involved in a hectic period of work that put further editing on the back burner. Only now as we approach Christmas have I had the space to revisit it. So here it is – attached.

We now have a short novel of correspondence between us. As I suspect we may not add much more to the debate as a result of this, my intention is to bring our three missives together as a paper for publication on my website. I do get enquiries from people who have had a successful track record of implementing RBA questioning the “Kittens” work, usually failing to reconcile their experience with your conclusions. It would be useful to refer them to our correspondence! Whilst I'm happy to add any further observations you might feel able to make, I fear I'm exhausted by the process now and can only urge you to consider my points carefully and recognise the contradictions. I absolutely do not want you to back pedal on your principal agendas of opposing Payment by Results and the Target Setting management. But your insistence on tarring RBA with the same brush is less than helpful and a distraction from the campaign in which we have a mutual interest. As I have said before, outcomes don't kill kittens, managers do!

I look forward to hearing from you in due course.

With good wishes,

David Burnby

Response (3) to an RBA Critique by Dr Toby Lowe

Commissioning Support Services for Ryan

I picked up on the Ryan case study example which provides a useful means of grounding the discussion and establishing principles. This also focusses our discussion on the performance accountability element of Results Based Accountability™ (RBA) in that we are examining accountability for a single service provider in terms of its relationship with a service user. Your view is that *“No organisation (or combination of organisations across a programme) can be accountable for Ryan's behaviour. There are too many external factors”*. I only partly agree with you here. If we are using public money to fund an intervention then it would be reckless in the extreme if we didn't hold someone accountable for the work we commission. But that said, I agree that one agency is in

itself most unlikely to bring about the desired outcomes (conditions of well-being) we require by working in isolation which is why partnership working is at the heart of the RBA process. We quite rightly hold a single entity accountable for performance, but in the expectation that success will only be achieved by working in partnership with others. If we abdicate responsibility, then everyone, and consequently no one is accountability which is unacceptable in the public realm.

It's important to recognise however that the RBA process does not rely wholly on outcome measures to evaluate and manage performance; data is just one part of a seven stage intuitive process:

1. Who are our customers (or clients/service users). In other words, who is the intervention for? In the Ryan example this is obvious though in more complex scenarios, this isn't always the case.
2. How can we measure if our customers are better off? What would success for Ryan look like? In this case, he would be outside of the criminal justice system, living in stable accommodation and would have addressed his drug misuse.
3. How can we measure if we are delivering services well? We are unlikely to achieve our customer outcomes by delivering poor quality services. You suggest that a key factor in Ryan's rehabilitation was the quality of the relationship he developed with his support worker. This is a quality measure.
4. How are we doing on the most important of these measures? This is about getting under the skin of the problems; understanding the forces and drivers that impact on Ryan and continually assessing what works, and what doesn't work.
5. Who are the partners that have a role to play in doing better? This acknowledges that the agency working alone cannot bring about the desired change in a scenario of any complexity. Ryan's girlfriend was clearly a key partner in this case. I'm sure there were others.
6. What works to do better? This is our plan for Ryan. It's based on our understanding learned from question 4 with the input and active engagement of partners
7. What do we propose to do?

These seven questions form the basis for a simple report card and it is of course an iterative process subject to constant review assessing what is working and what isn't. This report card would provide the reporting basis between the commissioner and the provider. If I was the commissioner, I'd be asking these seven questions to ensure that:

- Improving outcomes for Ryan was central to the provider's service provision
- That there is evidence of quality provision
- The provider is working in partnership with others, particular those closest to Ryan
- That a range of interventions are being used and assessed

This to me represents responsible commissioning and appropriate accountability. (The seven stage process incidentally also provides a robust script for staff supervision).

Next Generation Contracting

Friedman has developed a commissioning model he calls 'Next Generation Contracting' which incorporates this process. The model is designed to take us beyond the very crude performance

management methods which the public sector copied from the private sector. In 'First Generating Contracting', a purchaser would buy a given quantity of product or service and would probably specify quality standards. Whilst this worked fine for purchasing office supplies or painting council houses, it's not sophisticated enough to embrace the more complex conditions of well-being (outcomes) we are seeking when purchasing services in social care.

Friedman contends that whilst traditional service purchasing methods work fine for buying quantity and quality of service, they break down when applied to buying customer outcomes for much the same reasons as you suggest; i.e. case mix differences and perverse incentives. So in RBA, rather than taking about the traditional 'purchaser-provider split' we advocate a 'performance partnership' where funders and service providers work together to maximise performance. Central to this relationship is the seven stage report card process referred to above, not least to surface ways in which the purchaser can support the provider in achieving the desired outcomes. I cannot claim that in next generation commissioning, a purchaser would continually pour money into a service that failed to deliver any tangible outcome and arguably, it wouldn't be the best use of resources if it did, particularly if they were funding a poor quality service. But in a performance partnership, success would be judged not by crude data in isolation, but by assessment based on the seven step report card. RBA most definitely does not advocate the setting of unrealistic performance targets as a means of incentivising better performance. Friedman covers this in his book, pouring scorn on the notion:

*"Success usually means point to point improvement, moving from where we are now to a new better place. Crime should go down. High school graduation should go up. Every measure has a success direction and success is moving to a new place in that direction. "We will be successful if crime is reduced by 20% in the next two years" "Our school will be successful if reading scores go up by 5% per year". The problem is that the world doesn't work that way. Most conditions we are trying to change have a history. They have a momentum. Often the trend has been headed in the wrong direction for a long time. Sometimes, the best you can do is slow down the rate at which things are getting worse before you can flatten the curve and turn it up in the right direction. Every time we define success as point to point improvement, it is a set up for failure. Two years later, if we haven't hit the new lower (or higher) level, our work is declared a failure, and we go on to the next fad."*⁹

You suggest that RBA isn't about learning but on this you are quite wrong. It is a process that advocates a continuous learning process; the concept of the Performance Partnership between commissioner and provider is designed to adapt services to learn from what works and what doesn't thereby improving outcomes. Whilst the use of data is central to the RBA methodology, it is only one component of reporting. Friedman talks about experiencing outcomes to help determine appropriate measures, but also recognises the value of recording accomplishments and anecdotes as part of the reporting and accountability process. In a simple scenario based around a single service user, data would be but a small part of the reporting picture (Ryan's either offending or he isn't). More relevant in this case is demonstrating quality of provision through the report card process so the commissioner can reach a conclusion as to whether the service is worth continuing investment in.

⁹ "Trying Hard is Not Good Enough", P58

Ultimately, accountability between purchasers and providers in more complex systems has to have a significant data element which is essential to determine if things are getting better or worse. How else would you know? Is it realistic, for example, to suggest that government would spend several hundred thousand pounds on a programme to rehabilitate offenders without checking how many service users re-offended? Is it credible to advocate a system of funding services to make people better off without using data to see if it's working? Would such a system drive service improvement?

Organisational Culture

In your response, you speak about the culture of organisations and it's this subject that interests me greatly. Your principle objection to RBA seems to be that any system designed to measure outcomes will automatically be abused and you suggest that game play and manipulation of data is the only rational behaviour that will result in an organisation that adopts such a system. Your hypothesis is that this is inevitable when success is measured by factors which are outside of the service providers' control. But realising client outcomes is not wholly outside of the control of the service provider! As I've said elsewhere, a service provider **working in isolation** is unlikely to achieve service user outcomes in services of any complexity. Which is why working in partnership should be one of the quality measures the provider is held accountable for. Prisons for example, (always assuming we contract them both to rehabilitate offenders as well as incarcerate them) will not address recidivism without working in partnership with other agencies outside of their walls. They need to work with training providers to address basic numeracy and literacy difficulties; with employers to address employment; with health providers to address substance and alcohol abuse etc. The effectiveness of those partnership relationships should be part of the reporting process between purchaser and provider.

Friedman makes this point when talking about the paradox of control and how service providers have the least control over the most important measures (i.e. service user outcomes). You miss his point completely when you suggest he's advocating holding people accountable for things they are unable to control. What he's saying is that service providers must take greater control over service user outcome measures by working with others in partnership (as per the above example). Teachers could easily refuse to be held accountable for failing attainment levels by citing the many external factors that impact on student performance. But the good teachers don't abdicate responsibility in this way. They work with employers and higher education providers to drive aspiration; foster good working relationships with parents; get involved in after school and extracurricular activities etc. thereby gaining greater control over their client outcomes.

You suggest that I am being unrealistic about what happens when people use outcomes data for commissioning purpose. I know what happens when they don't! I've had meetings with commissioners who have multi million pound budgets yet have no clue what they're buying in terms of outcomes. They are spending money buying services by quantity alone with no notion if they're making a difference to anyone or not. The assumption seems to be that providing service equals people better off. Even worse, I see managers and commissioners accountable to no one; supervising interventions which are damaging to people as their understanding of performance management is limited exclusively to quantity of service delivered.

I agree that it's not irrational for service providers to game the system if their future depends on outcome measures, yet such abuse of the funding relationship is mitigated against by the performance partnership I describe where data is just one part of the reporting relationship. And if people get involved in this business because they want to make a difference (on both the purchaser and provider side) it is something of an insult to their motivation to suggest game play and fraudulent reporting is inevitable. In my (not inconsiderable) experience I cannot deny it happens but this is due almost inevitably to eroded relationships or poor contracting practice. We use RBA as a basis of minimising that risk.

Population and Performance Accountability

I fear you've misunderstood my emphasis on accountability by interpreting my argument as *“results for groups of service users are acceptable as performance management tools in ways that results for broader populations are not”*. This is not what I'm saying. Outcome data (indicators) at the whole population level measure distance travelled towards **whole population** outcomes. Performance Measures measure the effectiveness of interventions on **service-user** or **client populations**. So the economic activity rate for example is an indicator of progress towards an outcome of economic prosperity for a whole population. Number of service users (clients) moving into stable employment is a performance measure for an intervention. It's the difference between means and ends.

The accountability for a strategy to move a given community (whole population) towards greater prosperity lies with partners and stakeholders to design a collection of interventions with a reasoned chance of success. The data (indicators) tells us how effective our strategy is. The accountability for the performance of the various interventions we are commissioning and in particular, the service user outcomes, lies with managers and/or commissioners. In other words, indicators help us answer the question “Are we doing the right things?” Performance measures help us answer the question “Are we doing the right things well”. Means and Ends. Nothing to do with acceptability of tools!

Targets

My definition of a target in this context is “a specific desired future level of achievement”. RBA operates on the basis of measuring performance by distance travelled; or ‘turning the curve’. Friedman explicitly counsels against setting point by point targets for many of the reasons you suggest. But this doesn't mean we should operate in the absence of data. People soon get disillusioned in their work if they can't see how they are making a difference. They want to see evidence that they are learning from their experience and continually improving. That's why they work in this field; they want to make a difference. This approach is only a target in the sense that we want to improve performance as much as possible and as quickly as possible. If we're monitoring impact data closely we should be able to tell if we're making a difference. An important element of RBA is the learning process from continuous evaluation and an understanding of the ‘Story behind the baseline’ or the context in which we're operating. It shouldn't necessarily follow that if we're not making progress on a performance curve, then the project funding is automatically cut, particularly if there is a radical change in the ‘Story behind the Baseline’. After all, it's in the purchasers' interest to improve the performance. In consultation with the purchaser (in the ‘performance partnership’ model) we use the data to inform a change in tack, improve delivery or sharpen partnership arrangements. But if our interventions continually fail to make any measurable

impact then yes; the funding is likely to be cut. We cannot justify pouring money into funding interventions that don't make a difference; this work is too important for that.

Friedman addresses target setting specifically:

*Setting targets has been especially problematic. There is a long history of setting unrealistic performance targets and then beating people up when they fail to meet them. This practice is based on the misguided belief that such targets, coupled with fear of punishment will lead to better performance. In practice, unrealistic targets detract from credibility. And fear turns out to be one of the worst ways for managers to motivate people. People working within this punishment culture will try to pick measures they look good on or set targets they can easily meet. The measures are rarely the most important ones and the targets are meaningless. The organisation appears to be practising performance accountability, but it's a waste of time"*¹⁰

It's also important to recognise that in managing performance, RBA does not rely wholly on a single outcome measure, but on a balanced set of three to five measures of quality of services as well as outcome measures (answering the questions "How well did we do it" and "Is anyone better off" respectively). These are used in the context of historical performance to demonstrate and credit incremental improvement. This is precisely to militate against your submission that measuring outcomes results in perverse behaviour. He writes in the context of comparing performance data for a given service against established standards:

When standards are used to stretch organisations towards higher performance, this should never be done without also using baselines to give credit for incremental improvement. Failing to do this can lead to dysfunctional behaviour. A hospital was being judged against standards for in-patient death rates by type of operation. They figured out how to beat the system. If a patient was about to die, they were discharged so it wouldn't show up in their statistics. This hospital looked great on in-patient deaths. It is not a hospital you would want to go to.

This illustrates another important point. If you manage your programme using only one measure, you can look good on that measure at the expense of something else. Maybe the hospital should have tracked a second measure about the well-being of patients 6 weeks after discharge. If the discharged patients were all dead, this would hint at a problem. The hospitals goal should have been to optimise performance simultaneously on in-patient death rates and post-discharge recovery.

The danger of perverse incentives can be found in just about any system. If you're a high school principal and want to improve the school's attendance rate, then kick out the trouble makers. The attendance rate will go up. If you are a social services director and want to bring down social caseloads, slow down child abuse investigations. The foster care rates will begin to decline.

*We are not trying to find one magic measure for a programme, but rather a balanced set of three to five measures taken from "How well did we do it" and "Is anyone better off". These can be used to give a 'checks and balances' view of programme performance. Managers can work towards improving performance on all of them simultaneously"*¹¹

¹⁰ "Trying Hard is Not Good Enough" P87

¹¹ Ibid P86

You create an example of an employment support programme for people who have experienced homelessness which gets 15% of its service users into employment (25 people in your example). So, the performance measure for 'quantity of service' is '166 people served'. The 'better off' measure (or service user outcome) is 'number (25) and percentage (15%) of service users moved into employment'. You suggest that to demonstrate good performance, the measure the following year would automatically need to increase from 15% of service users to 20%. If the contract was managed in this way, I would agree: it's a target that is neither fair nor useful. But in the RBA next generation commissioning model, such crude performance management is explicitly discouraged. Reporting would be made on a variety of measures and, critically, would take into account the 'Story behind the baseline'. So maybe, because of the changing economic conditions, simply standing still on that measure would be considered good performance. This wasn't of course the case with the A4E work programme contract which was judged exclusively on getting people into work. And A4E wasn't using the RBA performance partnership methodology. That doesn't mean that outcome focussed service provision and performance management is wrong per se; it means that a more elaborate process should be in place; which is why I contend that RBA should not be lumped together with crude 'payment by results' approaches that lack this level of sophistication.

In your second example you suggest: *"If you commission a service based on changes to service users' BMI, Blood Pressure and Alcohol Consumption, what you will get are programmes that are brilliant at delivering improved data for service users on BMI, Blood Pressure and Alcohol Consumption. Service deliverers will cherry pick clients that they know will show improvements in these measures, whilst rejecting those that they know will be more difficult to make show improvements. They will put their resources into activities which reduce BMI, Blood Pressure and Alcohol Consumption, irrespective of whether that is what clients actually need. They will attempt to reclassify people to demonstrate improvement – what if we moved the boundaries of alcohol measurement? Finally, if all else fails, they will make the figures up"*

My first observation is that if you improve service users' BMI, blood pressure and alcohol they will be better off, so you are delivering service user outcomes, and the data will reflect this. Service providers may be tempted to cherry pick clients that are more inclined to want to address their unhealthy lifestyle, which is why, in next generation commissioning, good practice would specify under "How much did we do" specific demographic or service user profiles that must be addressed. Classification of measures will be negotiated between the commissioner and service provider and should only be changed following agreement on both sides. And whilst there are well documented cases of service providers rigging the figures (usually when unrealistic and/or unfair targets have been set in a crude commissioning agreement), such fraudulent behaviour is not the norm.

When training with RBA, and I address this issue, I use the analogy of RBA being like a knife. Knives can be life saving devices in the hands of skilled people when there is the right motivation. But the same tool that a surgeon may use for a lifesaving procedure can also have devastating consequences in unskilled hands (even with the right motivation). If the same tool is used as a weapon (i.e. the wrong motivation) it can be lethal. But we don't condemn knives per se. We attempt to create an environment conducive to fostering the right motivations and give people the skills to use the tools effectively. Of course, RBA can be damaging to people if it's being used incorrectly (such as is profiled in the Keevers et al paper: right motivation, unskilled application) and can be used fraudulently by people with the wrong motivation. So we use RBA as means of building on the right

motivation (my experience suggests that the vast majority of people working in this sector actually want to make a difference) and providing the skills to enable people to focus on outcomes without being hamstrung by complex processes or driven by crude commissioning targets.

Logic and Complexity

You suggest that my positioning of Friedman's view on Logic Models is contradicted by a piece on the same subject on his website. The piece you quote was written over ten years ago and does not (he tells me) reflect his current position on logic models. Friedman's current position is better reflected in the article "What's Wrong with Logic Models"¹² written for the Local Community Services Association (Australia) by Philip L. Lee, President of the Results Leadership Group¹³. Friedman promotes this article on his website¹⁴ which also features a parody of logic models.

The Philip L. Lee article contains a key paragraph which is central to your critique of RBA:

*"In the logic model world one thing leads to another in predictable or highly probable ways that can be written down in the form of a flow chart. Most people know this isn't the way things actually work. Causality is sometimes clear in simple systems. You hit a ball with a bat and the ball flies through the air in a parabolic arc. Hitting the ball with the bat causes it to fly through the air in a predictable way. But what causes a young person to drop out of high school? What is the right sequence of events that goes from giving advice to a young person to that young person deciding to stay in school? Things start to get complicated. Dropping out of school has many causes that can't be easily untangled. Logic models to the rescue. A flow diagram will show how specific program actions address all the causes that lead to school success. A little thought shows that logic models are incapable of representing the real world of uncertain causal relationships"*¹⁵

And this brings me to the apparent contradiction you cite between the "working backwards from means to end" approach RBA advocates and complexity theory. You say: *The problem with this is that it is impossible. You cannot "work backwards" in complex systems. Where results are the emergent properties of complex systems, 'working backwards' simply cannot be done. Which leaves Friedman with a problem – either he properly embraces complexity, or he rejects the idea of "working backwards" from ends to means.*

This is, I suggest a wholly false premise. Logic Models (for example) attempt to predict a 'logical' link from one stage of a process to another. **RBA doesn't work like this!** Complexity theory tells us this is difficult, if not impossible because of the infinite number and complexity of patterns of behaviour that will emerge from any given course of action. RBA attempts to identify these patterns as they emerge and work with service providers to adapt actions and interventions to suit. It's an iterative

¹² <http://www.lcsansw.org.au/documents/item/210>

¹³ Results Leadership Group "is the world's leading Results-Based Accountability™ resource and creator of the Results Scorecard™ software, works with governments and nonprofits internationally to achieve measurable improvements for communities and customers. Based in America, the company's affiliate office in Sydney is Results Leadership Group: Australia. Mark Friedman is a member of the Results Leadership Group.

¹⁴ http://www.resultsaccountability.com/new_stuff.htm

¹⁵ "What's Wrong with Logic Models" Philip L. Lee (Occasional Paper 1 published by the Local Community Services Association

and learning process. RBA recognises the theory of the “Adjacent Possible”; the notion that we have little idea precisely what the cause and effect result will be from any given intervention, but we can learn what is possible by embarking on the journey and observing the patterns of behaviour that emerge.

RBA is virtually unique amongst strategic planning processes in that it starts with the outcome by answering the question “What conditions of wellbeing do we want for our communities?” This is based on the belief that it is far better to move towards an outcome than it is to move away from a problem. Reaching consensus on where we want to be (which isn’t usually too difficult – there’s a remarkable consensus around whole population outcomes) creates the common ground necessary to bring partners and stakeholders together to move towards the best possible actions or interventions. When we are clear about what our outcome would look like if it were possible to achieve (by defining the appropriate indicators) we can then begin to try and understand the forces and drivers at work. We do this by working with the broadest possible cross section of partners and stakeholders. Views of residents and service users are given the same weight as technical research. We call this “The Story behind the Baseline”. Only when we have the richest possible story with the broadest possible cross section of partners and stakeholders can we hope to come up the best possible actions and interventions.

This is what we mean by working backwards from ends to means. RBA sees a strategic plan (or a Delivery Plan) as a moveable feast, something that is revisited on a regular basis as our understanding of the “Story behind the Baseline” changes as a result of the journeys we embark on. The reporting scorecards used in RBA are not just focussed on data: they are designed to continually monitor and review the Story and adapt actions and interventions to suit. The ‘next generation commissioning’ approach advocates a ‘performance partnership’ between commissioners and service providers in order to learn from this understanding and adapt contracts to suit. So I contend that RBA is very much better able to cope with complexity than most models in that it attempts to recognise emerging and changing patterns and how they impact on the system; adapting our practice to suit.

Keevers et al

I think you must have misunderstood my previous critique of the Keevers Article. I hadn’t suggested that the organisation featured in her research wasn’t using RBA. I suggested that the practitioners were using RBA wrongly and that Keevers has erroneously blamed the bad workmanship on the tools. Incidentally, your brief communication with Keevers merely serves to demonstrate her complete misunderstanding of what RBA is: she refers to “*Friedman’s version of Outcome Based Accountability*”/RBA. There is no other version of Outcome Based Accountability/RBA! As I indicated in my first paper to you, it is a trade-marked expression which refers to a very specific process and a collection of quite unique principles as presented in the book “*Trying Hard is Not Good Enough*” (and various academic papers published by the Fiscal Policy Studies Unit). The only conclusion one can draw from the Keevers paper is that if RBA is misinterpreted and/or misunderstood (which is clearly the case in her example), then it can distort good practice.

As far as I am aware, the Keevers et al paper is your sole point of reference that links RBA (as a specific process) to your wider agenda on outcomes-based performance management. I stick

doggedly to my view (and have explained fully why) this is a flawed piece of sloppy research that does nothing to support your wider argument. RBA (OBA as it is known in England) has been adopted as a thinking process in forty states of the USA and by hundreds of organisations in eight other countries across the world. Does it not occur to you that despite hundreds of applications of RBA/OBA across the world, you could only manage to dig up a single negative critique?

Where Next?

We have between us now a short novel of correspondence about RBA. I wish you continuing success in your campaign to discredit crude target setting performance management regimes and payments by results contracts. As I have stated before, I share your concerns and use every available opportunity in my training and consultancy practice to discredit and discourage these approaches. My wish is that you make the effort to understand RBA better and cease to link it as part of your wider agenda to the kind of practices I condemn. You will find a rigorous defence of RBA by practitioners in most places you speak where the process has been adopted and this will serve only as a distraction to your wider and more important message. There is a risk that by misrepresenting RBA (and/or quoting those who do) you will hinder a movement that is in fact an ally to your wider agenda.