

SYSTEMS CONCEPTS
IN EVALUATION
An Expert Anthology

EDITORS

Bob Williams
Iraj Imam

INTRODUCTION FROM THE SERIES EDITOR

It's a pleasure to welcome readers to the first volume in a new series sponsored by the American Evaluation Association. The aim of the series is to make high quality work in evaluation available at a modest price that will make it easier for people in the United States and other countries to become aware of and keep up with this relatively new and fast-developing field. The Monograph Series was conceived of and will mainly consist of relatively brief single-author works, but will deviate from that model when the occasion arises. As it happened, an unusual opportunity made it possible to inaugurate the series with this very timely and well-staffed anthology.

The series is overseen by the Publications Committee of the AEA, and supported by the Board of AEA, and was made possible as well as much improved by them, along with considerable help from the professional staff of the Association—most notably its Executive Director, Susan Kistler. This particular work also benefited greatly from its instigation and funding by the W. K. Kellogg Foundation, whose director of evaluation, Teri Behrens, was very helpful throughout; and from the extremely valuable comments of the two distinguished reviewers for AEA, Elliot Stern and William Trochim. The book content was designed by Rose Miller; it is being published for AEA through EdgePress of Inverness, California, whose indispensable Jeri Jacobson manages the logistics.

Hard copies of this publication can be obtained from EdgePress PO Box 69 Point Reyes CA 94956, or Amazon (<http://www.amazon.com>) for \$US18 soft cover and \$36 hard cover.

Your reactions to this book, and suggestions about other possible titles for the series would of course be much appreciated; after all, evaluators should practice what they preach!

Michael Scriven
scriven@aol.com

CONTENTS

ACKNOWLEDGEMENTS	I
INTRODUCTION	3
Systems Thinking for Evaluation <i>Gerald Midgley</i>	II
A Systemic Evaluation of an Agricultural Development: A Focus on the Worldview Challenge <i>Richard Bawden</i>	35
System Dynamics-based Computer Simulations and Evaluation <i>Daniel D Burke</i>	47
A Cybernetic Evaluation of Organizational Information Systems <i>Dale Fitch, Ph.D.</i>	6I
Soft Systems in a Hardening World: Evaluating Urban Regeneration <i>Kate Attenborough</i>	75
Using Dialectic Soft Systems Methodology as an Ongoing Self-evaluation Process for a Singapore Railway Service Provider <i>Dr Boon Hou Tay & Mr Bobby, Kee Pong Lim</i>	89
Evaluation Based on Critical Systems Heuristics <i>Martin Reynolds</i>	IOI
Human Systems Dynamics: Complexity-based Approach to a Complex Evaluation <i>Glenda H Eoyang, Ph.D.</i>	123
Evaluating Farm and Food Systems in the US <i>Kenneth A Meter</i>	14I
Systemic Evaluation in the Field of Regional Development <i>Richard Hummelbrunner</i>	16I
Evaluation in Complex Governance Arenas: the Potential of Large System Action Research <i>Danny Burns</i>	18I
Evolutionary and Behavioral Characteristics of Systems <i>Jay Forrest</i>	197
CONCLUDING COMMENTS	2II
BIOGRAPHIES	215



ACKNOWLEDGEMENTS

Producing this volume has been rather like running a medium-sized business. Indeed it would be a good systems case study on its own. We have tried to keep a collaborative ethos right through, so consequently the list of acknowledgements is extensive.

Our thanks to the following folk and no doubt others we have forgotten; you know who you are.

Robin Miller had the original idea. Michael Scriven and the American Evaluation Association provided the vehicle. Craig Russon made the link between the AEA and the Kellogg Foundation. Teri Behrens and the W K Kellogg Foundation supplied the grants that made the endeavour possible. The staff of the Center for Applied Local Research (CAL Research), especially Carol Coley, provided constant support.

The authors went way beyond the call of duty many times, and so did those whose potential contributions we didn't have space to include. The team of reviewers from all over the world looked at their allocated papers at least twice; Meenakshi Sankar, Leslie Goodyear, Hassan Qudrat-Ullah, Lee Mizell, Jacob Silver Olaronke Ladipo, Samruddhi Thaker, Thomas Chapel, Gene Lyle, Shannon Lee, Chris High, Graham Smith, Bridget Roberts, Alison Amos, Bill Harris, Bob Dick, Cheyanne Church, Eileen Franco, Greg Yelland, John Sherman, John Smith, Mel Tremper, Mike Lieber, Lindsey Poole, Rainer Loidl-Keil, Robin Miller, Aisha Shah, Tracy Henderson, Tony Berkley, Melissa Weenink, and Yee Sue Lin. Our apologies if we have missed someone.

CAL Research with help from Andie Williams of San Francisco Bay Area Evaluators (SFBAE) organized a two-day meeting between the authors, editors, and evaluators Beverley Parsons, Susan Hanson, Amy LaGoy, Tony Berkeley, and John Gargani. Ruhel Boparai took the notes. San Francisco Bay Area Evaluators (SFBAE) also hosted a dinner that allowed a whole bunch of SFBAE members to contribute to this publication.

Bill Trochim and Elliot Stern reviewed the whole draft and raised the standard even higher. Derek Cabrera, Patricia Rogers, Doug Fraser, Tessie Catsambas, and Michael Patton provided vital comments and support. Contributors to the Eval-Syst discussion group kept the endeavor grounded. Rose Miller did the design and layout



INTRODUCTION

Iraj Imam, Amy LaGoy, Bob Williams¹

Here is a short history of an idea.

We are somewhere in the 1960s. The idea is an approach to describing and assessing a situation in a way that can inform appropriate choices for action and improvement. The idea gains political support and the grudging acceptance of the academic world. Some academics criticise it for lacking an overarching theory, but reluctantly allow it into their departments and schools, partly because it is popular and lucrative.

Over time a network of consultants cluster around the idea, responding to the demands of clients who become interested in it.

The idea becomes fashionable in many parts of the world. It moves from being an idea to a field of inquiry to, some would argue, a trans-discipline like statistics. But the idea is applied differently in each part of the world.

There are arguments and splits. There are debilitating debates around about what is “real” and what is “perceived,” and the use of qualitative and quantitative data. These are partly resolved by those who argue the wise use of both. But the bruises persist. There are also debates over the role of the emerging field in empowering people, especially the disenfranchised. Over time the focus on a single stakeholder view of the situation is expanded to allow multiple stakeholder views. This emerging field is beset with communication difficulties and misunderstandings; the field has become so diverse that despite many attempts it defies a single definition or description.

Is this a description of the evaluation field or the systems field? The answer is both.

An intriguing aspect of the systems and evaluation fields is that they share many experiences, concepts, goals, even attitudes, yet know relatively little about each other. What each understands about the other is often crude and partial. Ask an evaluator about the systems field and you will probably hear about models, interconnections, and “holism,” but little more. Ask a systems practitioner about evaluation and you will hear about measurement, targets, and outcomes, and little else. These perceptions reflect only a small part of the two fields’ range of activity. They also fail to get to the core of what either field is about.

Despite drawing on some of the same philosophical, sociological, and scientific developments in the latter 20th century, the two fields have operated virtually independently since their inceptions. In recent years, however, some systems practitioners have begun applying systems thinking to evaluation work.

¹ With contributions from this volume’s authors, as well as Tony Berkley, Derek Cabrera, Tessie Catsambas, Susan Hanson, Beverly Parsons, and Bill Trochim.

And today there is growing interest among evaluators in what the systems field can offer them.

During a lively session at the 2005 Joint American Evaluation Association and Canadian Evaluation Society conference, participants described what was exciting about the use of systems concepts in evaluation:

- Makes you think differently.
- Offers more effective ways of dealing with complexity and complex situations.
- Links the local and global, across silos, sectors and disciplines.
- Provides tools to work with different opinions of stakeholders.
- Develops new ways of understanding situations.
- Pays attention to coalitions.
- Pays attention to properties that emerge unexpectedly.
- Puts multiple projects and topics into comparable forms.
- Acknowledges the richness and interdependence of real life.
- Helps identify leverage points; the differences that make a difference to a program and signal where best to intervene.
- Allows for measuring or accounting for dynamic changes in a program or system.
- Provides practical guidelines for using theory-of-change techniques.
- Recognizes the evolutionary nature of programs.

The evaluators at that session clearly understood that systems concepts had something to offer their evaluation work. To help deepen that understanding, this publication addresses three questions:

1. What key systems concepts do evaluators need to know?
2. How can evaluation benefit from using systems concepts?
3. What do evaluations based on systems concepts look like?

What key systems concepts do evaluators need to know ?

Let's conduct a thought experiment to show why this is such a challenging question. Replace the word "system" with "evaluation." Now imagine yourself an evaluator answering this question in front of an audience unfamiliar with evaluation. It's not an easy task. Now imagine a panel of other expert evaluators answering the question. The audience would get individually cogent answers, but would come away unsure whether there was a collective coherence: was there a single definition of evaluation, a coherent theory for making judgements about a situation, or examining the worth of a project or program or even for defining what we mean by project or program?

Despite this apparent incoherence, the evaluation field has a pretty clear sense of identity. The field has a boundary fence. Practitioners have an intuitive sense of what lies "inside" and "outside" that fence. Here is one of the great similarities between evaluation and systems, between being evaluative and being systemic: in both fields an overall understanding of the field is better achieved by observing

collective practice than by studying collective theory or agreed definitions. This doesn't mean that theory has no place in systems or evaluation, just that it is often more coherent when applied to specific methodologies than to the entire fields. So whilst it is possible to be clear about empowerment evaluation, utilization-focused evaluation, realist evaluation, soft systems, complex adaptive systems, and cybernetics, we are invariably hazy about evaluation and systems concepts in general.

As in the evaluation field, within the systems field common definitions were relatively easy to achieve when the field was new and compact. In the late 1970s Ackoff (1972) easily identified 32 core systems concepts. Broader intellectual traditions influenced and expanded the systems field during subsequent years. By 2003, Charles François' *Encyclopaedia of Systems and Cybernetics* had over 1700 entries. Schwartz's concept map of systems methods displays over 1000 items (Cabrera 2006).

This poses significant problems for volumes like this. How much of this broad field can we cover in 80,000 words? What should we cover? The boundary we chose is wider than some, and narrower than others would draw. This volume goes beyond variations of system dynamics (probably the most frequently documented group of systems concepts). Systems ideas and practice developed in different ways in different parts of the world and we wanted that diversity to be reflected in this volume. The authors hail from Australia, Austria, England, New Zealand, Singapore, the United States, and Wales. On the other hand we needed to erect a barrier somewhere and thus we chose not to explore methodologies used in systems inquiries (eg network analysis, knowledge management), that originated outside the systems intellectual tradition.

Gerald Midgley's chapter describes this intellectual tradition, and its parallels with the development of evaluation thought. Essentially he identifies three phases or "waves" of thinking about systems. Each "wave" or phase emerged in response to critical evaluations of the logic and methods of its predecessors. In the first phase, the focus was on improving a situation by *describing* in a more fundamental way the physical systems "out there" in the real world. In the second phase, the focus of attention shifted towards using systems concepts as tools to *understand* the real world in a more profound way. In other words, treating a situation as if it were a system. Or as Danny Burns states in his chapter, the systems described were not regarded as representations of reality (as in the first phase), but as mental constructions to enable learning. These constructions were usually built from exploring and considering multiple perspectives. The third "wave" or phase acknowledged that, in reality, not all perspectives are born equal. So for a truly systemic analysis and solution, each perspective is subjected to a critique that challenges the power structures and claimed expertise that gave it status.

Each phase or wave added to rather than replaced previous waves. Midgley likens this to pebbles from sequential waves building up on a beach. With such an accumulation of ideas broadening the field, universal statements about the nature of systems and systems inquiry are difficult to formulate and agree

upon. The evaluation field experienced some of this same loss of definition as it grew to encompass a greater variety of approaches. Where the expansion in methodological variety in both evaluation and systems has enhanced their ability to address the complexity of the world in which they operate – a net good, it has also made it difficult to define precisely and concisely what each field is.

So for those of you looking for coherence about what we consider to be relevant systems concepts for evaluation, our advice when reading this publication is to look for patterns rather than definitions. For us, three patterns stand out:

1. Perspectives. Using systems concepts assumes that people will benefit from looking at their world differently. For systems practitioners, this motivation is explicit, deliberate, and is fundamental to their approach. However, just looking at the “bigger picture,” or exploring interconnections does not make an inquiry “systemic.” What makes it systemic is how you look at the picture, big or small, and explore interconnections. A “system” is as much an “idea” about the real world as a physical description of it.
2. Boundaries. Boundaries drive how we “see” systems. Boundaries define who or what lies inside and what lies outside of a particular inquiry. Boundaries delineate or identify important differences (ie what is “in” and what is “out.”) Boundaries determine who or what will benefit from a particular inquiry and who or what might suffer. Boundaries are fundamentally about values – they are judgements about worth. Defining boundaries is an essential part of systems work/inquiry/thinking.
3. Entangled systems. One can observe and perceive systems within systems, systems overlapping other systems, and systems tangled up in other systems. Thus it is unwise to focus on one view or definition of a system without examining its relationship with another system. Where does one system begin and the other end? Is there overlap? Who is best situated to experience or be affected by that overlap? What systems exist within systems and where do they lead? A systems thinker always looks inside, outside, beside, and between the readily identified systems boundary. He or she then critique and if necessary changes that initial choice of boundary.

These three concepts are essential both for understanding systems-based interventions and for distinguishing them from other approaches to complex situations. They underpin the models, metaphors, methodologies, and methods used by the systems field. They provide the key to unlock the potential benefits of systems approaches to evaluation.

How can evaluation benefit from using systems concepts?

In the 12 chapters, you will see examples of systems-based approaches being used in evaluation when:

- there was a need to get to the core of the issue
- there was a need to get to the “undiscussibles” in an evaluation (eg Kate

- Attenborough, Martin Reynolds)
- there was a risk of getting lost in the complexity of the situation (eg Glenda Eoyang, Daniel D Burke)
 - assessing information needs across a program (eg Dale Fitch)
 - understanding the dynamics of a situation: Why do things emerge the way they do? (eg Daniel D Burke, Glenda Eoyang)
 - how do we define “value” or “worth” and at what cost to whom (eg Martin Reynolds, Ken Meter)
 - the program aims were not addressing the needs of the community being served (eg Kate Attenborough, Ken Meter)
 - it was important to raise the issue of those “unrepresented” in an evaluation (eg Martin Reynolds, Danny Burns)
 - simplifying the complexity of the situation risked missing the really important issue (eg Richard Hummelbrunner)
 - understanding the nature (and breadth) of the program impact was important (eg Danny Burns, Richard Bawden)
 - unravelling the interconnected strands of a complex situation (eg Jay Forrest)
 - Resolving multiple perspectives and multiple stated and unstated goals (Martin Reynolds, Glenda Eoyang, Kate Attenborough, Ken Meter)
 - there was a need to expose the assumptions underpinning a program or situation (eg Jay Forrest, Glenda Eoyang)
 - knowing where and how to place the boundary of the evaluation was critical (eg Martin Reynolds, Richard Bawden)
 - it was difficult to accommodate the differences amongst stakeholders (eg Boon Hou Tay).

These are areas evaluation often finds difficult, and it is in resolving these types of evaluation challenges that systems concepts will be most useful.

Systems concepts will be particularly useful to evaluators in situations where rigorous rethinking, reframing, and unpacking complex realities and assumptions are required. As one of the authors, Jay Forrest, puts it: “Systems approaches are for addressing untidy problems that don’t lend themselves to tidy packaging. Systems approaches offer a method for cutting through the mess of problem definition and surfacing a cleaner, sharper understanding of a situation, which may enable a clean useful answer. While systems approaches may find a clean answer, the reality is that complex, controversial problems rarely have clean, simple answers. Systems based approaches offer a possibility of packaging messy solutions in a clear manner that facilitates comprehension, understanding, and proactive action.”

Systems inquiry and evaluation thus tend to emphasise different understandings of both the task of inquiry and the situation under study. Even when it poses the same questions as an evaluation, an inquiry using the systems

concepts described in this volume is likely to interpret the answer from different perspectives. These include seeing the complicated as simple but not simplistic; being highly critical of boundaries that define what is “in” and what is “out” of the frame of inquiry, and the notion that deeper meaning-making is more likely to promote valuable action than better data. Bob Flood calls this “learning within the unknowable.” (Flood 1999)

What do evaluations based on systems concepts look like?

Evaluations influenced by systems concepts used in this volume are likely to generate rich yet simple descriptions of complex interconnected situations, based on multiple perspectives that ...

- ... build stakeholdings in the situation being addressed;
- ... are believed by stakeholders;
- ... help stakeholders build deeper meanings and understandings of the situation;
- ... reveal, explore, and challenge boundary judgements, that can inform choices for action by those who can improve or sustain a situation.

Let’s unpack this.

Simple descriptions of complex interconnected situations. Most people use maps, models and metaphors to describe a rich reality. In systems work, richness implies that the whole is can only be understood as a product of its parts plus the dynamic relationship between those parts. It is axiomatic in a systems-influenced evaluation that you may not be able to assess the whole from its parts, nor easily deduce the parts by observing the whole.

Many assume this implies that systems approaches to evaluation have to include every component of that situation plus its context, plus its environment. In fact, the implication is the opposite. Including everything in a systems inquiry doesn’t necessarily provide any deeper insights about the parts, nor does it necessarily offer more insights into the whole. Instead a systems-based approach to evaluation is concerned with what can be reasonably left out of the inquiry. It would, however, be deeply and openly aware of the consequences.

So a systems-influenced evaluation, no matter how comprehensive, remains incomplete. Importantly however, it will acknowledge and investigate that incompleteness. The richness of a systems inquiry is not about detail but about value. And the value is contained in the relevance of the inquiry to those affected by it. This is why we believe that systems-influenced approaches to evaluation have to build stakeholdings in the situations being addressed. Stakeholders are not passive players or mere informants – they are actively engaged in the critique about boundaries.²

2 Some of the so-called “hard” systems approaches do not inherently place as much emphasis on this. However, those that use these approaches frequently do – an example of how third phase ideas have influenced ideas that originated during the first phase.

It is critical for us that the description or rich reality of the situation be believed by stakeholders and build in them deeper meanings and understandings of the situation. Systems-influenced evaluation makes stakeholders aware of their beliefs. What we believe to be true defines our choices; it sets boundaries for our thoughts, understandings, and actions. Systems-influenced evaluation deliberately exposes our (evaluators' and stakeholders') assumptions about what is valid knowledge. This exposure comes through our embrace of *multiple perspectives*, a critical systems concept. Including multiple perspectives forces us to ask the questions: Whose reality is being considered? Who is defining what the situation is? Where do the limits of that situation lie?

Using systems concepts in evaluation helps us answer these questions by *revealing, exploring, and challenging boundary judgments* associated with a situation of interest. Decisions and insights about who or what is “in” and what is “out” of an inquiry, its boundaries, are key features of a systems inquiry. These boundary decisions cannot be accepted without question; nor are they contextual issues that once acknowledged can be ignored. In a systems-influenced evaluation, we believe the boundary is always in view and always up for debate, from the initial design stage to the end. Thus depictions of reality (models, maps, narratives, etc) are seen as dynamic, open-ended, and are designed to deal with puzzles, paradoxes, conflicts, and contradictions. Dialectic as well as dialogue is often a key feature in systems-influenced evaluations. Evaluators using systems concepts seek interconnections among disparate things. They define boundaries, cross and break boundaries, and create new ones – all in a manner relevant to the situation.

What can inform choices for action by those who can improve or sustain a situation depends in social power relationships and the epistemologies available. And whose epistemologies are privileged. Systems-influenced evaluation consciously explores ways to increase the voices of those in the margin in order that theirs might counterbalance the voices of the powerful. *Multiple perspectives* often reveal multiple, divergent, disputed purposes. Finding an appropriate course for action generates a need for explicit ways of addressing issues of coercion and power relations. The inquiry itself as well as the knowledge the use of multiple perspectives generates can create new alternatives for reform in the situation (or program.) These options will, of course, be better for some and worse for others. We believe that in a systems influenced evaluation the beneficiaries and victims, whether people or ideas, must be more than just acknowledged. Efforts need to be undertaken to mitigate the weaknesses of, and threats to those likely to be victimized or harmed by any change or reform.

Finally, does size matter? One common misconception is that systems approaches are more appropriate for large scale, longer-term evaluations. But like evaluation, systems approaches are scaleable. They can be applied to big and small issues, over short and long timeframes – the key is the *nature* of the situation not the *scale* of the situation. The examples in this volume tend to be on the medium

to large size, and the constraints on the evaluator fairly loose – but we have used systems approaches in very small and constrained situations. They can be used to evaluate entire programs, or to design a questionnaire. Nothing is too big or too small.

References

- Ackoff, R.L. 1971. Towards a system of systems concepts. *Management Science*, 17: 11.
- Cabrera, D. 2006. *Systems Thinking*. Dissertation, Cornell University, Ithaca.
- Flood, R. 1999. *Rethinking the Fifth Discipline; Learning within the Unknowable*. Routledge.
- François, C. 2003. *International Encyclopedia of Systems and Cybernetics*. 2nd edition. Germany: KG Saur.
- Schwartz, E. 2000. *Some Streams of Systems Thought*. [Online] Available from: <<http://www.igss.net/gPICT.jpg>>

How to read this Monograph

Read Gerald Midgley’s chapter first, since it will give you a broad overview of the systems world and its development over the past half century. Gerald’s chapter provides the foundation for the remaining chapters. You can read the rest of this Monograph in any order. However, early chapters tend to focus on single methods, later ones on multiple methods. The Monograph ends with some broader applications of systemic thought.

Gerald Midgley’s opening chapter explores the origins of systems thought and lays out its historical and intellectual development. His three “waves” of systemic thought provides an important framework for what follows. **Richard Bawden** partners Gerald Midgley’s opener, by discussing the implications of the three “waves” of systemic thought in helping reshape the lives of Indian farmers, plus a final challenge for evaluators, or anyone, setting off down the systems road. Next are chapters by **Dan Burke**, **Dale Fitch**, **Kate Attenborough**, **Boon Hou Tay**, **Martin Reynold**, and **Glenda Eoyang** that describe tools used by common systems approaches; system dynamics, cybernetics, soft systems, complex adaptive systems, and critical systems thinking. **Ken Meter’s** chapter then shows how you can combine these tools. **Richard Hummelbrunner** adds further tools to the box with his insights into considering the evaluation task as a system, and **Danny Burns** describes the close relationship between systems thinking and action research. **Jay Forrest’s** chapter lays out a framework for working in a broader systemic manner. **Bob Williams**, **Iraj Imam**, and **Amy LaGoy** return to reflect on what all this might mean for the wider systems and evaluation canons.

Systems Thinking for Evaluation

Gerald Midgley

In this pivotal chapter, Gerald achieves several things at the same time. He weaves together the intellectual development of the systems field, how this has influenced practice and importantly the relevance of all this to evaluators and evaluation. In doing so he sets the boundaries of this volume; what systems concepts and methods are covered here and which ones are not. It provides a vital map by which you can navigate and place yourself within both the systems and evaluation territories. You will return to this chapter often.

1. Introduction

This introductory chapter is founded on the assumption that, while evaluators already have an interesting armoury of both quantitative and qualitative methods at their disposal, they can still enhance their theory and practice with reference to systems thinking. The term “systems thinking” is commonly used as an umbrella term to refer to approaches that seek to be more holistic than those scientific (and other) methodologies that concentrate attention on a relatively narrow set of predefined variables. Although systems thinkers *aspire* to holism, most nevertheless appreciate that it is impossible for human thought to be all encompassing (eg Churchman 1970; Bunge 1977; Ulrich 1983), and they also recognise that narrowly focused studies to answer well-defined questions can still usefully inform systemic analyses (eg Checkland 1985). Many researchers and practitioners have therefore developed systems approaches that not only offer their own insights, but also contextualise or complement those to be gained from uses of other methods and methodologies (eg Midgley 2000).

In the context of evaluation, it is widely recognised (eg Zammuto 1982) that setting narrowly defined goals for a service or organisation, and measuring the achievement of these alone, may result in the evaluator missing positive benefits that lie outside the scope of the evaluation. Indeed, perverse behaviours may be generated in the service or organisation being evaluated. Recently, I encountered a good example in a UK hospital ward that prided itself on its speedy through-put of patients being given endoscopies (investigations of the oesophagus or stomach via a camera passed through the mouth). The staff claimed to be one of the top performing units in the country. However, they met (and exceeded) their targets by refusing to use anaesthetics, routinely used by similar services elsewhere, unless patients absolutely insisted. The result was exceptional through-put, because the staff had eliminated the time it takes for people to recover from the anaesthetic, but the price was paid in patient pain and discomfort.

While this sort of problem has long been recognised by evaluators, and solutions (in the form of multivariate evaluations and/or the inclusion of multiple stakeholder perspectives) have been devised, there is still considerable scope to

enhance both theory and practice when it comes to making evaluations more holistic. The systems literature is a good source of ideas for these enhancements because, as we shall see in more detail shortly, it encompasses almost a century of research into what holism means (although, as M'Pherson, 1974, points out, the roots of these ideas can be traced back much further). In the last hundred years, systems approaches have evolved considerably, and many of the paradigm shifts that are familiar to evaluators (see, for example, Guba and Lincoln 1989) have parallels in the systems literature. While the diversity of systems paradigms and methodologies can be confusing for newcomers to the field (necessitating introductory chapters like this one), the focus on holism (albeit understood in different ways in the different paradigms) has remained constant.

In my view, because of the diversity mentioned above, it would not be productive to start the main body of this chapter by defining terms. While it is possible to characterise systems thinking very broadly as a holistic way of thinking, and there are certainly common concepts used in many of the systems paradigms represented in the literature (Marchal 1975), as soon as we dig beneath the surface the meanings of these concepts are inevitably contested, or what is left unsaid by focusing only on a limited set of words becomes too important to ignore. Arguably, a better way to introduce the diversity of systems thinking is to explain the emergence over time of some quite different approaches. It will become apparent as my description of these different systems perspectives unfolds that the movement has been in a continual process of development and change, and that many practical problems of direct concern to evaluators have been addressed along the way.

Of course, I will only be able to provide an indication of the diversity of systems approaches, given almost one hundred years of innovation. Because of this, it is important for me to explain the focus I have chosen and what I have consciously excluded from my review. In presenting successive developments of systems ideas I use a "wave" metaphor. A wave throws useful materials onto the beach, and these are then added to and sometimes rearranged when the next wave hits. I argue that there have been three waves of systems research since the 1940s, each of which offers a different basic understanding of systems and consequently a different methodological approach.

Inevitably, all metaphors highlight some features of a situation while hiding others (Morgan 1986). In this case, the fact that some researchers continue to develop older ideas in useful directions even after new waves come along is made less visible than I might like by the wave metaphor. Nevertheless, the advantage of using this metaphor is that it focuses attention on some of the major shifts in understanding that have taken place, leaving us with a wide range of systems approaches to learn from.

My review of the first wave of systems thinking starts with a brief explanation of its early 20th-century origins, and then moves on to review the three major, interrelated lines of research that characterised systems thinking in the 1940s and 1950s: general system theory (GST), cybernetics and complexity science. The

implications for evaluation of these early (and a few relevant later) ideas will be considered. When discussing the second and third waves, however, I will narrow the focus to what are now commonly called “management systems” approaches. There is a massive volume of literature on second and third wave systems approaches related to ecology, philosophy, family therapy, technology and so many other domains of inquiry that it is impossible to list them all. I will concentrate on management systems because researchers in this area have produced many methodologies and methods to enhance the holistic focus of policy making, planning, decision making and (on occasion) evaluation. Therefore, this is arguably the most relevant specialism within systems thinking for evaluators. However, even with a narrowed focus on management systems, it’s still not possible to cover more than a fraction of the relevant and useful ideas in the literature. I therefore recommend that the interested reader use this chapter as a starting point for his or her own explorations. There are also a number of books of edited readings out there, the most recent (and arguably the most comprehensive) being a four-volume set covering a much broader set of ideas than this short paper (Midgley 2003a).

2. The First Wave of Systems Thinking

While it is quite possible to trace ideas about holistic thinking and change processes back to the ancient Greeks, especially Heraclitus (Crowe 1996) and Aristotle (M’Pherson 1974), systems thinking as we know it today is widely recognised to be a child of the first half of the 20th Century. However, the birth of this child is shrouded in controversy. This is because, for many years in the West, people attributed the earliest articulation of systems ideas to von Bertalanffy (1950, 1956, 1962, 1968), who started lecturing and writing in the 1930s on what he called “general system theory”. It was only after several decades had passed that it came to light that a very similar set of ideas had been produced by Bogdanov in Russia in 1910–1913, before von Bertalanffy started writing. There seems to be no hard evidence that von Bertalanffy knew of Bogdanov’s work prior to producing his own, but comparisons of the writings of the two authors have generated some interesting discussions about their similarities and differences (Gorelik, 1987; Dudley, 1996).

Although there is controversy surrounding the origins of contemporary systems thinking, it is nevertheless clear that, immediately post-WWII, three parallel and mutually supportive fields of inquiry came to prominence: general system theory (GST), cybernetics and complexity science. The common themes running through these fields of inquiry, and their influence on science, technology and public planning after the war, helped their advocates to launch the first wave of systems thinking and establish a new systems research community that is active to the present day (and is now growing more rapidly than ever). Below, short summaries of GST, cybernetics and complexity science are provided, and their implications for evaluation are discussed.

2.1 GENERAL SYSTEM THEORY

General system theory (GST) gained significant attention in the mid-20th century because it proposed an antidote to what some people call “reductionist” science. This is the kind of science that continually divides its subject matter into more and more specialist disciplines; focuses on small numbers of linear, causal relationships between phenomena; and looks for explanations of all phenomena in terms of their smallest identifiable parts. GST proposes that we can transcend the boundaries of narrow, specialised disciplines by looking at things as *open systems*. Fundamentally, an open system exchanges matter and energy with its environment (Koehler, 1938; von Bertalanffy 1950; Kremyanskiy 1958). It is a unity made up of *organised elements*. This organisation is crucial because it gives rise to properties of the system (called “emergent properties”) that cannot be found in a disorganised collection of the same elements. An example is a person who can only remain alive as long as his or her parts are organised in a set of particular relationships with one another. A random heap of organs is not a living human being. GST proposes that systems of all kinds share certain common characteristics that can be described through the use of mathematics as well as ordinary language. Therefore, by studying systems as *general phenomena*, and by working out the laws that they all obey, we can learn more about the functioning of *specific systems* (eg cells, organisms, families, organisations, communities, societies, planets, and galaxies). Key writers on GST include von Bertalanffy (1956, 1968) and Boulding (1956), and it is worth going back to Bogdanov (1910–1913) too.

There are several useful concepts that evaluators can take from GST, one of which is the idea that it is possible to evaluate the *viability* of an organisation – ie how well it performs as an open system seeking to survive in a turbulent environment, and whether it exhibits the necessary characteristics to thrive. While a number of writers have translated the open system idea from GST into approaches of relevance to planning and management (eg Trist and Bamforth, 1951; Trist et al 1963; Emery and Trist, 1965; Emery and Thorsrud, 1969, 1976; Kast and Rosenzweig 1972; Emery 1993), a key author from an evaluation perspective is Beer (1959, 1966, 1981, 1984, 1985). He realised that it is possible to derive indicators of viability by looking at how all higher order open systems (including organisms with complex nervous systems as well as wider social systems) need to be organised in order to survive, and he combined this idea from GST with cybernetic thinking (see below) to produce his viable system model (VSM). The VSM proposes that, for an organisation to become and remain viable in a complex and rapidly changing environment, it must carry out five essential functions (see Beer 1981, 1985, for details). According to the VSM, the key to effective organisation is not only to make sure that all five functions exist, but also to ensure that communications between the functions are appropriate and effective. Together, these functions manage the information and decision flows necessary for effective organisation. The model can be used to design new organisations or to diagnose current organisational failings. In the latter mode it is quite explicitly

an evaluation approach. Also see Hronek and Bleich (2002) for another GST-inspired approach to evaluating the effectiveness of organisations and services.

Another useful concept for evaluation that can be derived from GST is the idea that open systems are structured in hierarchies: systems are embedded within wider systems (eg individual human beings live within communities which are embedded in ecosystems). The whole can enable and/or constrain the parts, and the parts can contribute to and/or challenge the stability of the whole. This way of thinking can help us understand why social systems are so often resistant to change: they can self-organise in order to maintain a relatively stable state. One implication for evaluation of thinking in terms of hierarchies is that both the environment and sub-systems of any particular “system in focus” might be important if we want to understand what is going on. A useful rule of thumb might be: once the system in focus has been defined (ie a particular policy implementation system, service or organisation has been chosen for evaluation), look at least one level up and one level down to gain a greater understanding of what is happening. For an example of an evaluation that is explicit in its use of systemic hierarchy theory, see Benkö and Sarvimäki (2000).

2.2 CYBERNETICS

Also in the mid-20th century, cybernetics took shape. The basic idea in cybernetics is *feedback*, and a classic example of a simple mechanical feedback system is a thermostat controlling a set of radiators. In a cold room, the thermostat will switch on the radiators. Then, as the room heats up, it will receive feedback from its thermometer and will turn the radiators off. Therefore, through feedback, it is able to control the temperature of the room. In different kinds of cybernetic systems, feedback may have another function: instead of being used to maintain a relatively stable state, it may prompt the system to move in a particular direction without there being a balancing constraint to return it to its original starting point. There has been a lot of communication between cyberneticians and systems theorists because feedback is used by all kinds of systems to monitor and respond to their environments. Both ideas are therefore strongly related to one another. Key writers in the early days of cybernetics include Weiner (1948), Ashby (1956), and Bateson (1972).

A key insight from both GST and cybernetics is that researchers cannot actually step outside social and ecological systems when they observe and communicate their observations (von Bertalanffy 1968; Bateson 1970; von Foerster 1979; Maturana 1988): observation is situated within particular relationships, and it influences those relationships through communications and actions. The implication for evaluation is that it is better to recognise and take explicit responsibility for our involvements in situations than to pretend that our systems practice is a-contextual and value-neutral (Churchman 1970; Ulrich 1983; Alrøe 2000; Midgley 2000, 2003b). I recognise that the idea of non-neutrality can be problematic for those evaluators who are commissioned to provide summative

assessments of services, products or organisational processes: commissioners often value the independence of the evaluator. It also has to be acknowledged that the principle of non-neutrality is more strongly implied in the works of later, more philosophical cyberneticians (eg Bateson 1972; von Foerster 1979; Morgan 1982; von Glasersfeld 1985; Maturana 1988) compared with the mathematical writings of earlier cyberneticians with an engineering bent (eg Weiner 1948; Shannon 1948; Shannon and Weaver 1949). However, to me this issue is not black and white. It is possible to acknowledge that value judgements flow into decisions about what criteria to evaluate against, what to include in or exclude from an evaluation, etc, while still preserving a degree of institutional independence (ie there is some separation of the commissioner and evaluator through their different institutional affiliations, even though both may participate in the same wider ecological, cultural and linguistic systems). Also, the use of techniques of self-reflection and stakeholder dialogue can make the values flowing into an evaluation more transparent than they would be if the evaluator pretended to genuine neutrality – and, to a degree, stakeholder dialogues can “democratise” these values too.

Another cybernetic insight of potential importance to evaluators is that, without the use of methodological aids, human consciousness has a tendency to over-simplify causality, making many feedback loops and interdependencies invisible. Bateson (1979) gives a good example when he talks about how people commonly explain causality in biological evolution. It is said that, because many trees in Africa only have leaves high up in the air, giraffes had to grow longer necks to reach them (causality is unidirectional in this explanation). However, it would be equally possible to claim that the trees evolved their shape in response to the nibbling of giraffes! This is another unidirectional explanation that makes as much (or as little) sense as the first one. In preference to either of these, it might be useful to see giraffes and trees in a dynamic pattern of interaction characterised by feedback over time.

There are at least two significant implications for evaluation in this insight. The first is that it can be useful to model feedback processes as an aid to understanding. There are several systems methodologies offering mathematical modelling tools, including systems engineering (eg Hall 1962; Jenkins 1969), an approach focusing on the design of whole organisational systems, using quantitative methods, to meet given purposes in an optimal manner; and systems analysis (Quade and Boucher 1968; Optner 1973; Quade et al 1978; Miser and Quade 1985, 1988), a methodology for assessing costs, effectiveness and risk given multiple scenarios. However, one of the most widely used in support of evaluation and organisational learning, and one of the most sophisticated in terms of representing feedback, is system dynamics (eg Forrester, 1961, 1969; Roberts et al, 1983; Senge, 1990; Morecroft and Sterman, 1994; Sterman, 1994; Vennix, 1996; Maani and Cavana, 2000).

System dynamics gives evaluators a useful tool to model complex feedback processes in a manner that can help to make transparent why certain effects

might have occurred (or might occur in the future). Particular hypotheses about key interactions can be proposed, and then a model can be built to see whether the events that actually transpired can be simulated based on these hypotheses. Through iterations of generating and testing hypotheses, learning about successes and failures in the system of concern can be facilitated. Furthermore, Morecroft (1985) argues that system dynamics can be used to evaluate how sets of seemingly optimal actions at the sub-system level can combine to produce sub-optimal, or even catastrophic, outcomes at the system (organisational) level: local actors with inevitably limited knowledge make what seem to them to be rational decisions, but the interaction of these locally rational decisions in the context of the whole organisation may create unexpected effects. The production of these effects can be modelled using system dynamics to demonstrate to the sub-system participants that their actions need to change.

The second implication of the complex nature of causal relationships highlighted through research into cybernetics is that, because human thoughts and behaviours are part of wider systems of causality, evaluators need to be very careful about engaging in “blame games”. This is particularly important when evaluations are focused on allocating responsibility for failures: eg airline crashes or medical accidents. It may be the case that a doctor “made a mistake”, but the context of this mistake may have been that she was at the end of a thirty six hour shift without sleep; that departmental politics made relationships in the medical team strained; and that, due to financial stringencies, three doctors were trying to cover a case load that would normally be the responsibility of six. To what extent should the doctor be held accountable for the wider organisational and policy dysfunctions that put her in a risky situation? The cybernetic insight that causality has a systemic character, together with the insight from GST that organisations can usefully be described as open systems, has given rise to the *systems failure* evaluation approach (eg Fortune and Peters 1990, 1995). This offers a methodology and model for evaluating organisational failures without simply attributing blame to a single individual. While there may be cases where gross negligence or malicious intent can reasonably be attributed to one person despite wider systemic factors, it is nevertheless the case that most failure evaluations will need to look at causality more broadly if they are to avoid scapegoating “easy targets”, and if lessons are to be learned about how similar failures can be avoided in future.

It is not always easy to gain acceptance for this kind of evaluation in a culture characterised by a worldview where linear causality predominates, and where the free choice of an individual is accepted as the logical starting point of a causal chain. This worldview inevitably gives rise to a desire to blame (Douglas 1992). Nevertheless, some significant in-roads have been made in life-critical areas of human activity where it has been essential to enable the open reporting of mistakes so that learning can take place and future lives can be saved. A prime example is the area of medical error, where some organisations have stopped the “blame game” in order to prevent damaging cover-ups and encourage more systemic

evaluations than would have been conceivable just a decade previously (eg IOM Committee on Quality of Health Care in America 2000; Scarrow and White 2002).

2.3 COMPLEXITY SCIENCE

The third and final line of systems research initiated in the 1940s and 1950s is complexity science. As far as I am aware, Weaver (1948) was the earliest writer to identify complexity as a field of inquiry, and it was already clear in the late 1940s that when writers were discussing *organised* complexity, they were essentially talking about systems. Indeed, the terms “systems” and “complexity” have been very closely associated throughout the history of the systems enterprise – so much so that it is common to hear people say that the purpose of *applied* systems thinking is to help us deal with complexity (eg Flood and Carson 1993). A common understanding of complexity is that it results from the number of elements and/or interactions in a system being beyond the capacity of the observer to easily understand. However, even a small number of elements and/or interactions may be complex when they are in a dynamic and difficult to predict process of change. Also, the introduction of different normative or subjective perspectives into a situation can generate complexity when people have to come to terms with new ways of seeing that challenge their taken for granted assumptions (Flood, 1987; Midgley, 1992a). Interestingly, the 1990s witnessed a massive up-surge of both academic and popular interest in complexity science, using new mathematical ideas to explore the characteristics of complex systems. Key writers include Weaver (1948), Simon (1962), Prigogine (1987), Gell-Mann (1994) and Cilliers (1998).

An important insight for evaluation is that, while there is a good degree of regularity and order in the world, most systems (especially human, social and wider ecological systems) are nevertheless highly complex and therefore our ability to understand and predict their behaviour is limited. Sometimes, when a system is poised on the cusp of a change, the tiniest perturbation can send it spinning off in one direction rather than another (Prigogine, 1987). For this reason, evaluators should be wary of assuming “systemic invariance” (no change in the social or organisational status quo): most of us with experience of evaluation have engaged in projects with the good intention of bringing about much-needed improvements, only to find that a wider political or organisational initiative, or even a seemingly “minor” perturbation (such as the departure of a key champion for the project), has pulled the rug from underneath the evaluation. Complexity science provides some theory that helps explain such situations, and tells us to keep an eye out for changes to the context of an evaluation that may effect how it needs to proceed.

Other insights from complexity science have different methodological implications for evaluation. Allen (1988) discusses the phenomenon of *emergence*: new characteristics of complex systems emerge over time. This means that the evaluation criteria that seem most relevant today may become redundant, or might need to be supplemented by others, tomorrow. Allen also warns planners and managers to avoid commissioning grand modelling projects which purport

to make predictions of the future based on the assumption that a researcher (or evaluator) can account for every variable impacting on a system of concern. Not only is full comprehensiveness impossible at any given point in time, but the variables to be accounted for in a model need to be regularly rethought as decision makers encounter newly emergent properties of the system in focus. This means that relatively modest modelling projects, that adequately (rather than comprehensively) represent the history of a service or organisation for particular purposes, are generally more useful for enabling learning in evaluations than “super models” that take so long to build that they risk becoming redundant before they are ready to use (de Geus 1994; Sterman 1994).

A final insight of relevance to evaluation that emerges from complexity science actually problematises the neat conceptualisation of nested systems (systems having both sub-systems and supra-systems) mentioned in relation to GST. Relationships can sometimes be so complex that seeing things in terms of clearly differentiated levels can be counterproductive (an interaction with something defined as lying outside the immediate sub- and supra-systems may be highly important), so visualising a *network* of interactions can sometimes be more appropriate (Capra 1996; Taket and White 2004). However, in my view, seeing things in terms of *both* nested systems and networks can be beneficial. The rule of thumb I suggested earlier (look one level up and one level down from the system in focus) should not be operationalised mechanically. The idea of nested systems gives the evaluator a picture of where possible drivers for and constraints on the system in focus might lie, but keeping in mind the possibility that complex relationships may violate neat hierarchies sensitises one to the importance of looking out for crucial interactions that may not be represented in the usual focus on three levels of nesting.

Having considered the implications for evaluation of first wave systems thinking (in particular, GST, cybernetics, and complexity science), I will now proceed to discuss the second wave, narrowing the focus to *management* systems thinking.

3. The Second Wave of Systems Thinking

The first wave of systems thinking gained great popularity in the 1950s and 1960s. However, in the late 1960s (and even more so in the 1970s and early 1980s), significant questions began to be asked, both about the philosophical assumptions embodied in the first wave, and the consequences of its practical application.

3.1 CRITICISMS OF FIRST WAVE THINKING

The modelling approaches in the first wave were criticised for regarding models as representations of reality rather than as aids for the development of inter-subjective understanding (Churchman 1970; Checkland 1981a; Espejo and Harnden 1989; de Geus 1994). As some of the interveners who used these approaches believed that they had unique insights into the nature of complex systems, granted by the use of their systems methods, they set themselves up (and

were regarded by others) as experts in systems thinking. Thus, they fell into the trap of making recommendations for change without properly involving those who would be affected by, or would have to implement, that change. The result could often be recommendations that were regarded as unacceptable by stakeholders, and were therefore not implemented, or were resisted if implementation was attempted (Rosenhead 1989).

There was also a tendency in the 1960s for systems scientists, especially those working on public policy modelling for local authorities in the USA, to try to build generic “super models”: representations of whole cities and regions with masses of interacting variables rather than more modest models for specific purposes. The dream of truly integrated planning motivated these modellers, but they were mostly defeated by the scale and complexity of the task they set themselves (Lee 1973). This is an important lesson for evaluators, and is a mistake that hopefully will not be repeated given the greater knowledge that we now have of complexity and the inevitable limits it imposes on human understanding (Cilliers 1998), even when aided by models.

The first wave modelling approaches were also criticised for viewing human beings as objects that could be manipulated as parts of larger systems, instead of individuals with their own goals which may or may not harmonise with wider organisational priorities (Checkland 1981a; Lleras 1995). In consequence, several authors pointed out that the first wave of systems thinking failed to see the value of bringing the subjective and inter-subjective insights of stakeholders into activities of planning and decision making (Churchman 1970; Ackoff 1981; Checkland 1981a; Eden et al 1983).

Furthermore, it has been argued that most of the first wave systems approaches assume that the goal of the person or organisation commissioning an evaluation or systems analysis is unproblematic, when it is actually common to find that goals are unclear or there are multiple viewpoints on which goal it is most appropriate to pursue (Checkland 1981a; Jackson and Keys 1984). In such circumstances, to take a cynical view, it is relatively easy for the commissioner to manipulate applications of the first wave modelling approaches: unless the intervener has a strong sense of ethics, and some understanding of participatory practice, use of these approaches may support only the views of the person providing the money, allowing the opinions of others to be ignored (Jackson 1991).

Finally, some of the first wave approaches have been criticised for inheriting the emphasis of open systems theory (eg von Bertalanffy 1968) on organisms and organisations adapting to survive in ever-changing environments. Thus, the vision, goals and objectives of the organisation are not generally open to critique on normative grounds: surviving and thriving in a turbulent environment is seen as an end in itself, rather than a means to deliver some other socially-valued ends (Ulrich 1981; Jackson 1991).

All these criticisms led to a significant paradigm shift in systems theory and its application. A second wave of systems thinking was born. In this new wave,

“systems” were no longer seen as real world entities, but as constructs to aid understanding. The emphasis was on dialogue; mutual appreciation; the intersubjective construction of understandings; and searching for accommodations between different perspectives. Arguably, the authors who are best known for generating this paradigm shift in management systems are Churchman (1970, 1979), Ackoff (1981) and Checkland (1981a), although many more authors than these actually contributed to the change. Over the coming pages I will briefly review Churchman’s fundamental contribution to rethinking the systems idea. As Ackoff and Checkland have both been more concerned with planning and management than with evaluation, I will pass over their contributions (but I recommend them to the interested reader).

3.2 RETHINKING THE SYSTEMS IDEA

Churchman’s (1968a,b, 1970, 1971, 1979) work is incredibly broad in its scope and is therefore difficult to summarise. Here, I have chosen to highlight what I consider to be his most important contributions, although others might disagree with my emphases. Churchman (1970) proposes that, if a change is to be justifiably called an improvement, then reflecting on the boundary of analysis is crucial. What is to be included or excluded is a vital consideration: something that appears to be acceptable or desirable given a narrowly defined boundary may not be seen as such at all if the boundaries are pushed out. For this reason, Churchman argues that as much information as possible should be “swept in” to analyses, allowing the most inclusive and therefore most ethical position to emerge – but without compromising practicality through a confusing *over-inclusion*.

In comparison with the earlier ideas of the general system theorists, this way of thinking involves a fundamental shift in our understanding of the nature of a system (and hence the meaning of “holism” from a second wave perspective is likewise transformed). Prior to the work of Churchman, many people assumed that the boundaries of a system are “given” by the structure of reality (eg the skin of the human body marks the boundary of the person as an organism or living system). In contrast, for Churchman, boundaries define the limits of the knowledge that is to be taken as pertinent in an analysis.

Also, when it comes to social systems, pushing out the boundaries of analysis may involve pushing out the boundaries of who may legitimately be considered a decision maker (Churchman, 1970). Thus, the business of setting boundaries defines both the knowledge to be considered pertinent and the people who generate that knowledge (and who also have a stake in the results of any attempts to evaluate or improve the system). This means that there are no “experts” in Churchman’s systems approach, at least in the traditional sense of expertise where all relevant knowledge is seen as emanating from just one group or class of people: widespread stakeholder participation is required, sweeping in a variety of relevant perspectives. Indeed, Churchman (1968a) coined the phrase “the systems approach begins when first you see the world through the eyes of another”

(p231). The implication for understanding “holism” is that, from a second wave perspective, it no longer means seeking a comprehensive exploration of the nature of a system. Rather it means expanding the boundaries of our knowledge about whatever we are studying, together with the boundaries defining the set of people involved in generating that knowledge.

Not only did Churchman introduce this fundamental change in our understanding of “system”, but he also discussed critique. In examining how improvement should be defined, Churchman (1979) followed Hegel (1807), who stressed the need for rigorous self-reflection, exposing our most cherished assumptions to the possibility of overthrow. To be as sure as we can that we are defining improvement adequately, we should, in the words of Churchman (1979), pursue a “dialectical process”: this involves seeking out the strongest possible “enemies” of our ideas and entering into rational argumentation with them. Only if we listen closely to their views and our arguments survive should we pursue the improvement.

Here we have several shifts in perspective with significant implications for evaluation practice (also see Ulrich, 1988, who has written in more detail about the relevance of Churchman’s thinking for evaluation). First, there is an *epistemological* shift (a shift from one theory of human knowledge to another) involved in the move from first to second wave systems thinking. While most first wave systems thinkers took it for granted that human knowledge reflects reality (with various degrees of error), second wave writers talked about how human beings *construct* social realities, either at the level of the individual through our selective observations (eg Maturana 1988; Maturana and Varela 1992), or inter-subjectively through the shared use of language (eg Maturana 1988) and methodologies (eg Checkland 1981b, 1985). This is not a “solipsist” view, that there is no reality outside the human mind. Rather, it emphasises the impossibility of absolute proof of what is real, and therefore suggests we shift the emphasis in inquiry from supposedly objective modelling (or “truth-seeking”) approaches to ones which encourage mutual appreciation of different perspectives (acknowledging multiple realities), shared learning, accommodations, and the critique of assumptions underpinning our own and others’ perspectives. In response to this shift, an evaluator might want to engage in activities of self-reflection (because s/he knows that his/her own perspective is partial and limited); communicate around criteria for evaluation (because it might be possible to accommodate multiple perspectives on relevant criteria); seek out qualitative data about different perspectives rather than assume that a given set of quantitative indicators is adequate; and adopt a formative methodology, because ongoing learning is viewed as being more appropriate than a single, summative judgement that merely reflects the social norms of the moment.

Another second wave shift with implications for evaluation was from expert-led modelling to the development of *participative* methodologies (eg Ackoff 1981; Checkland 1981a; Mason and Mitroff 1981). This was partly based on the new

understanding of holism, now meaning inclusion of multiple perspectives (as opposed to the comprehensive representation of a system), but was also based on the critique of first wave practice where the assumption that the researcher was an expert had caused significant problems in many projects (see section 3.1 above). A fundamental assumption of second wave participative methodologies is that people are more likely to take “ownership” of analyses (including evaluations), and thereby implement the recommendations arising from them, if they can play a part in defining the goals and remits of those analyses (and even, in many cases, carry them out themselves). This insight is common to some evaluation paradigms too (Guba and Lincoln 1989).

The final implication for evaluation that I will discuss here comes from Churchman’s (1979) advocacy of dialectical processes: seeking out the strongest possible “enemies” of our ideas and entering into rational argumentation with them. While many evaluation methodologies compare a service, product or organisation with something else (eg what happens without the service, or when a different service is provided), the technique of comparing the *idea* behind the service with the ideas of its opponents is less frequently used by evaluators. The advantage of doing the latter is that it subjects the basic assumptions (or boundaries of consideration) underpinning what is being evaluated to critical analysis. Simply comparing a service with a control, or with another service seeking to accomplish the same thing, does not surface the possibility that the purposes being pursued by the service are inadequate in the first place.

4. The Third Wave of Systems Thinking

Finally I wish to discuss a third wave of systems thinking, taking us from the 1980s to the present day. All the third wave ideas reviewed below have been produced under the banner of “critical systems thinking” (for some edited readings on this, see Flood and Jackson 1991a; Flood and Romm 1996). While the third wave is actually broader than critical systems thinking alone (see Midgley 2003a, for some other relevant works), I see critical systems approaches as being of particular value because (amongst other things) many of them provide a rationale for taking the best methods from both the first and second waves and harnessing them into a broader systems practice. This pragmatic focus on mixing methods seems to be common to both the systems and evaluation communities.

4.1 CRITIQUES OF THE SECOND WAVE OF SYSTEMS THINKING

In the late 1970s and early 1980s, several critiques of second wave systems thinking were launched, primarily on the grounds that the participative methodologies that characterised this wave did not account sufficiently for power relationships within interventions, and/or conflicts built into the structure of society (Thomas and Lockett 1979; Mingers 1980, 1984; Jackson 1982).

The criticism of the lack of attention paid to power relations stems from the observation that stakeholders do not always feel able to speak openly in front of

one another for fear of the consequences (Mingers 1980, 1984; Jackson, 1982). Thus, in some situations, second wave systems approaches are likely to reinforce the viewpoints being promoted by the holders of authority without necessarily accounting for voices that might have been (wittingly or unwittingly) silenced.

However, there was also a set of criticisms focused on conflicts built into the structure of society. One such criticism came from an explicitly Marxist position: Thomas and Lockett (1979) tried to draw out similarities between the Marxist agenda and second wave systems thinking, and commented on the absence of a theory of society in the latter. From a Marxist perspective, it is a problem that stakeholders can reach collaborative agreements through the use of second wave systems methods without necessarily changing their basic relationship in terms of the ownership and control of economic resources: it suggests that these systems methods can be used to facilitate a “false consciousness” amongst disadvantaged sectors of society, where they feel that they have achieved something worthwhile through dialogue without ever touching the inequalities that, from a Marxist perspective, are a fundamental cause of many social problems.

Other criticisms came from a non-Marxist, but still explicitly emancipatory, perspective. Mingers (1980, 1984) and Jackson (1982) suggested that the focus of second wave systems thinking on participation and dialogue is right, but a theory of emancipation is needed to enable second wave methods to be harnessed in the service of a more deeply beneficial rather than merely superficial social change. These authors argued that second wave systems methodologies are “regulative”: that is, evaluation and intervention usually has such a narrow focus that wider unjust political and economic relationships are taken as the environment of the system of concern, and therefore participants are encouraged to adapt their organisations and/or policies to these relationships instead of challenging them. An emancipatory approach would encourage people to question the political status quo when adaptive behaviours are judged to be insufficient, with a view to enabling broader action for social change.

Very soon after these criticisms began to surface in the literature, the second wave systems thinkers came under attack from a new direction. Some people had become concerned that the systems community was being torn apart by a paradigmatic war between first and second wave thinkers – and yet both forms of thinking are necessary to deal with different kinds of problem. In 1984, Jackson and Keys published what was to become a highly influential paper, arguing that the first and second waves should be regarded as complementary rather than in competition with one another, and that *methodological pluralism* should be welcomed. Jackson and Keys’ paper (and subsequent developments of the argument by many other authors, but especially Jackson 1987a,b) became one of two key foundation stones upon which the new, third wave of systems thinking was built. The other foundation stone was Ulrich’s (1983) methodology of critical systems heuristics, which focuses on how the motivations for, control of, expertise in and legitimacy of any social system design can be considered critically and systemically by those

who are involved in and affected by it. Below, both these foundation stones are reviewed, and their implications for evaluation are considered. Then my own work on systemic intervention is discussed, as this seeks to integrate the main themes from earlier third wave writings in order to offer a new methodological approach that can draw upon a wide range of systems methods (plus techniques from other traditions, including evaluation) in a flexible and responsive manner that is able to account for processes of marginalisation (tied up with power relations) and the values and concerns of multiple stakeholders.

4.2 CRITICAL SYSTEMS HEURISTICS

Ulrich (1983, 1987) takes Churchman's (1970, 1979) understanding of the importance of boundaries to systemic analysis (reviewed in section 3.2) in a new and challenging direction. Ulrich agrees that Churchman's desire to sweep the maximum amount of information into understandings of improvement is *theoretically* sound, but also acknowledges that the need to take practical action will inevitably limit the sweep-in process. He therefore poses the question, how can people rationally justify the boundaries they use? His answer is to develop a methodology, critical systems heuristics, which can be used to explore and justify boundaries through debate between stakeholders. Ulrich (1996) calls the process of exploration and justification "boundary critique".

An important aspect of Ulrich's (1983) thinking about boundaries is that boundary judgements and value judgements are intimately linked: the values adopted will direct the drawing of boundaries that define the knowledge accepted as pertinent. Similarly, the inevitable process of drawing boundaries constrains the values that can be pursued. Boundary critique is therefore an ethical process. Because of the focus on dialogue and debate between stakeholders in dealing with ethical issues, a priority for Ulrich (1983) is to evolve practical guidelines that planners and ordinary citizens can both use equally proficiently to conduct boundary critique. For this purpose, he offers a list of twelve questions which can be employed by those involved in and affected by planning (the involvement of those affected being vital) to interrogate what the system currently is and what it *ought* to be. These twelve questions cover four key areas of concern: motivation, control, expertise and legitimacy. In the view of Ulrich (1988), comparing the *is* with the *ought* in relation to a service, product or organisation provides an effective evaluation approach.

In my view, there is significant potential for using Ulrich's twelve questions in evaluations, not least because they cut to the heart of many issues that are of fundamental concern to people in communities who find themselves on the receiving end of policies and initiatives that they either do not agree with or experience as irrelevant. In my own research, I have used these questions with (amongst others) people with mental health problems recently released from prison (Cohen and Midgley 1994; Midgley 2000); older people in sheltered housing (Midgley et al 1997, 1998; Midgley 2000); and young people who have run away

from children's homes (Boyd et al 1999, 2004; Midgley 2000). Ulrich (1983) claims that his questions can be answered equally proficiently by ordinary people with no prior experience of planning or evaluation as they can by professionals, and my experience tells me that he is right – with the caveat that the questions should be made specific to the evaluation in hand, and also need to be expressed in plain English (the original questions contain some academic language that will be inaccessible to many people).

4.3 TOWARDS METHODOLOGICAL PLURALISM

Ulrich's (1983) work was launched fully-formed into the systems community, and had a gradual but ultimately substantial influence on third wave thinking. The other key argument kick-starting the third wave, that systems practitioners should embrace methodological pluralism and mix methods, took several more years to fully evolve. Building on the work of Jackson and Keys (1984), Jackson (1987a) argued that three different types of systems thinking are useful to deal with three different types of problem: first wave systems thinking is useful when there is agreement between stakeholders on the nature of the problem situation and the goals to be pursued; second wave thinking is useful when there is non-coercive disagreement between key players, and this requires debate and learning to find a way forward; and critical systems heuristics is useful in situations characterised by coercion, when there are barriers to debate between stakeholders but this can be improved by amplifying the voices of those who are marginalised or disadvantaged. This is actually an over-simplification of Jackson's (1987a) argument, but more details can be found elsewhere (Jackson, 1987a, 1991). Essentially, Jackson offered a framework that aligned different systems approaches with different problem contexts. Later, Flood and Jackson (1991b) went on to embed this within a new methodology for creatively exploring problematic situations; choosing an appropriate systems approach; and implementing it (also see Flood 1995, and Jackson 2003, for later developments of this methodology).

Whether or not the reader wants to use Jackson's (1987a) framework for aligning systems approaches with different problem contexts (Midgley, 1997, reviews a range of criticisms of it), the basic idea of methodological pluralism is very important if we are to develop a really flexible and responsive evaluation practice. The fact is that no methodology or method (whether it comes from the systems tradition or anywhere else) can do absolutely everything people might want. Therefore, being able to draw upon multiple methods from different paradigmatic sources can enhance the systems thinking resource we have available for evaluation and intervention (eg Jackson 1987b, 1991, 2000; Flood 1990; Gregory 1996; Mingers and Gill 1997; Midgley 2000).

The theory and practice of this idea applied to evaluation has been explored quite extensively in the systems literature. The earliest writings proposing contingency frameworks of systems (and other) evaluation methods, showing how different approaches might be applicable to different contexts, were published in

the late 1980s (Midgley 1988, 1989), and these ideas were then developed further in the early 1990s (Midgley 1990, 1996; Midgley and Floyd 1990; Gregory and Jackson 1992a,b; Gregory 1993). Later, the use of contingency frameworks was abandoned by most of these and other authors (see Taket and White 1997, and Midgley 2000, for reasons), but the focus on theoretically-informed methodological pluralism remained.

4.4 SYSTEMIC INTERVENTION

The final third wave perspective I want to cover is my own work on systemic intervention, which draws together and further develops the other third wave ideas reviewed above. I should note that this has been used to frame several evaluations (as well as planning and development projects) (eg Midgley 2000).

In relation to boundary critique, I argue that there is often more going on than just the inclusion or exclusion of stakeholders and issues: it is not uncommon to find particular groups and the issues that concern them becoming *marginalised* (Midgley 1992b, 2000) – neither fully included in nor excluded from the system, and subject to strong labelling and ritual treatment (I use the terms “sacred” and “profane” to indicate the potency of the valuing or devaluing that marginalised people and issues are subject to). I also offer a model of marginalisation processes, demonstrating how marginalisation comes about through conflicts between groups making different value and boundary judgements, and how it can become entrenched in institutional arrangements. I suggest that, to avoid naïve approaches to formulating remits for evaluations and designing participatory methods, evaluators need to take issues of marginalisation seriously. One way to systemically address such issues is through the design of methods that can support dialogue across boundaries; revalue the contributions of marginal groups; and sweep in issues that might not normally receive proper consideration (Boyd et al 2004).

I have also advocated methodological pluralism in my own work (Midgley 2000), but do not now offer a generic contingency framework of methods aligned with problem contexts. This is because, as I see it, all evaluation scenarios have their own unique features and are often “multi-layered” (Weil, 1998): the evaluator has to consider multiple possible impacts and relationships, involving participants, other stakeholders and his/her own team. In my experience, multi-layered scenarios usually require a creative mix or design of methods tailored to the specific situation. Contingency frameworks tend to force choices between discrete methods. I also have several other reasons for not using a generic framework that I will not go into here (see Midgley, 2000, for details). Instead, I argue for an ongoing learning process where the practitioner considers methods in terms of their stated purposes and principles; the theories that underpin them; the ideological assumptions they make; and the practical results they have achieved in previous interventions (Midgley 2000).

However, the main added value that comes from the systemic intervention approach is arguably the *synergy* of boundary critique (drawing together the ideas

of Churchman 1970, Ulrich 1983, and my own thinking about marginalisation) with methodological pluralism. If boundary critique is practiced on its own, it is possible to generate some very interesting social analyses, but there is a danger that these will not be acted upon unless other methods for systemic evaluation and planning are used too (Flood and Jackson 1991b). Also, there is a danger in practicing methodological pluralism without up-front boundary critique: this can give rise to superficial diagnoses of evaluation scenarios. If a complex issue is defined from only one limited perspective without reflecting on values and boundaries, and issues of marginalisation are neglected, then the outcome could be the use of a systemic evaluation approach that misses or even exacerbates significant social problems (Ulrich 1993). The synergy of boundary critique and methodological pluralism ensures that each aspect corrects the weaknesses of the other (Midgley 2000; Boyd et al 2004).

I suggest that this kind of systems approach is not only able to address issues of values, boundaries and marginalisation in defining complex evaluation scenarios, but it also has the potential to deliver all the utility of previous systems (and other) approaches because it explicitly advocates learning about, and drawing methods from, those approaches to deliver maximum flexibility and responsiveness in systemic evaluations.

5. Conclusion

In this paper I have surveyed a wide range of systems ideas, focusing in particular on those that I believe could be particularly useful for systemic evaluation. I have described three “waves” of systems thinking, the final one being about the need for boundary critique (reflections on and dialogue about values, boundaries and marginalisation) and methodological pluralism (drawing upon and mixing methods from a variety of different paradigmatic sources).

Now, one criticism that has been levelled against the idea of methodological pluralism is that drawing methods from all three waves of systems thinking as well as the evaluation literature is asking too much of researchers and practitioners who have not had the benefit of studying the full range of systems methodologies on degree programmes. Too much learning is required before the researcher can get started, and busy professionals with existing methodological preferences may be resistant to trying methods from other paradigms (Brocklesby 1995, 1997; Mingers and Brocklesby 1996). If you are reading this introductory chapter having had no previous experience of using systems approaches, and are feeling daunted by the prospect, then I would like to leave you with one key idea. Rather than think that you need to know the full range of systems methods before beginning to practice, *start from where you are now*. Begin with the systemic insights and methodological resources you already have and step outwards from there. Ask yourself if there are just one or two ideas or methods that might enhance your existing systemic awareness, and try synthesising these with what you already do. I argue that building systemic resources requires a theoretical and methodological learning

journey in relation to practice, and does not require a full body of knowledge to be assimilated in advance (Midgley 2000). However, learning can be considerably enhanced by engaging with others at the same stage of the journey, and with those who are already further down the road. In my view, systems thinking has the potential to make a significant difference to evaluation, so (if you have not already done so) I invite you to take the first steps on your own learning journey – and share your experiences with others so that the whole evaluation community can be enriched in the process.

References

- Ackoff, R L. 1981. *Creating the Corporate Future*. New York: Wiley.
- Allen, P M. 1988. Dynamic models of evolving systems. *System Dynamics Review*, 4: 109–130.
- Alrøe, H F. 2000. Science as systems learning: Some reflections on the cognitive and communicational aspects of science. *Cybernetics and Human Knowing*, 7: 57–78.
- Ashby, W R. 1956. *Introduction to Cybernetics*. Chichester: Wiley.
- Bateson, G. 1970. Form, substance, and difference. *General Semantics Bulletin*, 37: 5–13.
- Bateson, G. 1979. *Mind and Nature: A Necessary Unity*. London: Wildwood House.
- Bateson, G. 1972. *Steps to an Ecology of Mind*. Northvale NJ: Jason Aronson.
- Beer, S. 1959. *Cybernetics and Management*. Oxford: English Universities Press.
- Beer, S. 1966. *Decision and Control*. Chichester: Wiley.
- Beer, S. 1981. *Brain of the Firm*. 2nd ed. Chichester: Wiley.
- Beer, S. 1984. The viable system model: Its provenance, development, methodology and pathology. *Journal of the Operational Research Society*, 35: 7–25.
- Beer, S. 1985. *Diagnosing the System for Organisations*. Chichester: Wiley.
- Benkö, S S & Sarvimäki, A. 2000. Evaluation of patient-focused health care from a systems perspective. *Systems Research and Behavioral Science*, 17: 513–525.
- Bertalanffy, L von. 1950. The theory of open systems in physics and biology. *Science*, 111: 23–29.
- Bertalanffy, L von. 1956. General system theory. *General Systems*, 1: 1–10.
- Bertalanffy, L von. 1962. General system theory – A critical review. *General Systems*, 7: 1–20.
- Bertalanffy, L von. 1968. *General System Theory*. London: Penguin.
- Bogdanov, A A. 1913–1917. *Bogdanov's Tektology*. 1996 edition. Dudley, P (ed). Hull: Centre for Systems Studies Press.
- Boulding, K E. 1956. General systems theory – the skeleton of science. *Management Science*, 2: 197–208.
- Boyd, A, Brown, M, and Midgley, G. 1999. *Home and Away: Developing Services with Young People Missing from Home or Care*. Hull: Centre for Systems Studies.
- Boyd, A, Brown, M, and Midgley, G. 2004. Systemic intervention for community OR: Developing services with young people (under 16) living on the streets. In, *Community Operational Research: OR and Systems Thinking for Community Development*. Midgley, G and Ochoa-Arias, A E (eds). New York: Kluwer.
- Brocklesby, J. 1995. From single to multi-paradigm systems research. In, *Systems for Sustainability: People, Organizations, and Environments*. Stowell, F A, Ison, R L, Armon, R, Holloway, J, Jackson, S, and McRobb, S (eds). New York: Plenum.
- Brocklesby, J. 1997. Becoming multimethodology literate: An assessment of the cognitive difficulties of working across paradigms. In, *Multimethodology: The Theory and Practice of*

- Combining Management Science Methodologies. Mingers, J and Gill, A (eds). Chichester: Wiley.
- Bunge, M. 1977. General systems and holism. *General Systems*, 22: 87–90.
- Capra, F. 1996. *The Web of Life: A New Synthesis of Mind and Matter*. London: HarperCollins.
- Checkland, P. 1981a. Systems Thinking, Systems Practice. Chichester: Wiley.
- Checkland, P. (1981b). Rethinking a systems approach. *Journal of Applied Systems Analysis*, 8: 3–14.
- Checkland, P. 1985. From optimizing to learning: A development of systems thinking for the 1990s. *Journal of the Operational Research Society*, 36: 757–767.
- Churchman, C W. 1968a. *The Systems Approach*. New York: Dell.
- Churchman, C W. 1968b. *Challenge to Reason*. New York: McGraw-Hill.
- Churchman, C W. 1970. Operations research as a profession. *Management Science*, 17: B37-53.
- Churchman, C W. 1971. *The Design of Inquiring Systems*. New York, :Basic Books.
- Churchman, C W. 1979. *The Systems Approach and its Enemies*. New York: Basic Books.
- Cilliers, P. 1998. *Complexity and post-modernism: Understanding Complex Systems*. London: Routledge.
- Cohen, C and Midgley, G. 1994. *The North Humberside Diversion from Custody Project for Mentally Disordered Offenders: Research Report*. Hull, Centre for Systems Studies.
- Crowe, M. 1996. Heraclitus and information systems. *Systemist*, 18: 157–176.
- Douglas, M. 1992. *Risk and Blame: Essays in Cultural Theory*. London: Routledge.
- Dudley, P. 1996. Back to basics? Tektology and general systems theory (GST). *Systems Practice*, 9: 273–284.
- Eden, C, Jones, S, and Sims, D. 1983. *Messing About in Problems*. Oxford: Pergamon.
- Emery, F E. 1993. Characteristics of socio-technical systems. In, *The Social Engagement of Social Science, Volume II: The Socio-Technical Perspective*. Trist, E and Murray, H (eds). Philadelphia PA: University of Pennsylvania Press.
- Emery, F E and Trist, E L. 1965. The causal texture of organizational environments. *Human Relations*, 18: 21–32.
- Emery, F E and Thorsrud, E. 1969. *Form and Content in Industrial Democracy*. London: Tavistock.
- Emery, F E and Thorsrud, E. 1976. *Democracy at Work*. Leiden: Nijhoff.
- Espejo, R and Harnden, R J. 1989. The VSM: An ongoing conversation... In, *The Viable System Model: Interpretations and Applications of Stafford Beer's VSM*. Espejo, R and Harnden, R J (eds). Chichester: Wiley.
- Flood, R L. 1987. Complexity: A definition by construction of a conceptual framework. *Systems Research*, 4: 177–185.
- Flood, R L. 1990. *Liberating Systems Theory*. New York: Plenum Press.
- Flood, R L. 1995. *Solving Problem Solving*. Chichester: Wiley.
- Flood, R L and Carson, E R. 1993. *Dealing with Complexity: An Introduction to the Theory and Application of Systems Science*. 2nd ed. New York: Plenum.
- Flood, R L and Jackson, M C (eds). 1991a. *Critical Systems Thinking: Directed Readings*. Chichester, Wiley.
- Flood, R L and Jackson, M C. 1991b. *Creative Problem Solving: Total Systems Intervention*. Chichester: Wiley.
- Flood, R L and Romm, N R A (eds). 1996. *Critical Systems Thinking: Current Research and Practice*. New York: Plenum.
- Foerster, H von 1979. Cybernetics of cybernetics. In, *Communication and Control in Society*. Krippendorf, K. (ed). New York: Gordon and Breach.
- Forrester, J W. 1961. *Industrial Dynamics*. Cambridge MA: MIT Press.
- Forrester, J W. 1971. Counterintuitive behavior of social systems. *Theory and Decision*, 2:

- 109–140.
- Fortune, J and Peters, G. 1990. The formal system paradigm for studying failures. *Technology Analysis and Strategic Management*, 2: 383–390.
- Fortune, J and Peters, G. 1995. *Learning from Failure: The Systems Approach*. Chichester: Wiley.
- Gell-Mann, M. 1994. Complex adaptive systems. In, *Complexity: Metaphors, Models, and Reality*. Cowan, G, Pines, D, and Meltzer, D (eds). Reading MA: Addison-Wesley.
- Geus, A P de 1994. Modeling to predict or to learn? In, *Modeling for Learning Organizations*. Morecroft, J D W and Sternman, J D (eds). Portland OR: Productivity Press.
- Glaserfeld, E von. 1985. Reconstructing the concept of knowledge. *Archives de Psychologie*, 53: 91–101.
- Gorelik, G. 1987. Bogdanov's tektologia, general systems theory, and cybernetics. *Cybernetics and Systems*, 18: 157–175.
- Gregory, A J. 1993. *Organisational Evaluation: A Complementarist Approach*. Ph.D. Thesis. Hull, University of Hull.
- Gregory, A J and Jackson, M C. 1992a. Evaluating organizations: A systems and contingency approach. *Systems Practice*, 5: 37–60.
- Gregory, A J and Jackson, M C. 1992b. Evaluation methodologies: A system for use. *Journal of the Operational Research Society*, 43: 19–28.
- Gregory, W J. 1996. Discordant pluralism: A new strategy for critical systems thinking? *Systems Practice*, 9: 605–625.
- Guba, E G and Lincoln, Y S. 1989. *Fourth Generation Evaluation*. London: Sage.
- Hall, A D. 1962. *A Methodology for Systems Engineering*. Princeton: Van Nostrand.
- Hegel, G W F. 1807. *The Phenomenology of Mind*. 2nd edition in English, 1931. London: George Allen and Unwin.
- Hronek, C and Bleich, M R. 2002. "The less-than-perfect medication system": A systems approach to improvement. *Journal of Nursing Care Quality*, 16(4): 17–22.
- IOM Committee on Quality of Health Care in America. 2000. *To Err is Human: Building a Safer Health Care System*. Washington D: National Academy Press.
- Jackson, M C. 1982. The nature of soft systems thinking: The work of Churchman, Ackoff and Checkland. *Journal of Applied Systems Analysis*, 9: 17–29.
- Jackson, M C. 1987a. New directions in management science, In, *New Directions in Management Science*. Jackson, M C and Keys, P (eds). Aldershot: Gower.
- Jackson, M C. 1987b. Present positions and future prospects in management science. *Omega*, 15: 455–466.
- Jackson, M C. 1991. *Systems Methodology for the Management Sciences*. New York: Plenum.
- Jackson, M C. 2000. *Systems Approaches to Management*. New York: Kluwer/Plenum.
- Jackson, M C. 2003. *Systems Thinking: Creative Holism for Managers*. Chichester: Wiley.
- Jackson, M C. and Keys, P 1984. Towards a system of systems methodologies. *Journal of the Operational Research Society*, 35: 473–486.
- Jenkins, G. 1969. The systems approach. *Journal of Systems Engineering*, 1: 3–49.
- Kast, F E and Rosenzweig, J E. 1972. General systems theory: Applications for organization and management. *Academy of Management Journal*, December 1972: 447–465.
- Koehler, W. 1938. *The Place of Values in the World of Fact*. New York: Liveright.
- Kremyanskiy, V I. 1958. Certain peculiarities of organisms as a "system" from the point of view of physics, cybernetics and biology. English trans, 1960. *General Systems*, 5: 221–230.
- Lee, D B. 1973. Requiem for large-scale models. *Journal of the American Institute of Planners*, 39: 163–178.
- Lleras, E. 1995. Towards a methodology for organisational intervention in Colombian

- enterprises. *Systems Practice*, 8: 169–182.
- Maani, K and Cavana, R. (2000). *Systems Thinking and Modelling: Understanding Change and Complexity*. Auckland: Prentice Hall.
- Marchal, J H. 1975. On the concept of a system. *Philosophy of Science*, 42: 448–468.
- Mason, R O and Mitroff, I I. 1981. *Challenging Strategic Planning Assumptions*. New York: Wiley.
- Maturana, H 1988. Reality: The search for objectivity or the quest for a compelling argument. *Irish Journal of Psychology*, 9: 25–82.
- Maturana, H R and Varela, F J. 1992. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Revised ed. Boston MA: Shambhala.
- Midgley, G. 1988. *A Systems Analysis and Evaluation of Microjob – A Vocational Rehabilitation and Information Technology Training Centre for People with Disabilities*. M.Phil. Thesis. London, City University.
- Midgley, G. 1989. Critical systems: The theory and practice of partitioning methodologies. *Proceedings of the 33rd Annual Meeting of the International Society for General Systems Research (Volume II)*, held in Edinburgh, Scotland, on 2–7 July 1989.
- Midgley, G. 1990. Creative methodology design. *Systemist*, 12: 108–113.
- Midgley, G. 1992a. Pluralism and the legitimation of systems science. *Systems Practice*, 5: 147–172.
- Midgley, G. 1992b. The sacred and profane in critical systems thinking. *Systems Practice*, 5: 5–16.
- Midgley, G. 1996. Evaluation and change in service systems for people with disabilities: A critical systems perspective. *Evaluation*, 2: 67–84.
- Midgley, G. 1997. Mixing methods: Developing systemic intervention. In, *Multimethodology: The Theory and Practice of Combining Management Science Methodologies*. Mingers, J and Gill, A (eds). Chichester: Wiley.
- Midgley, G. 2000. *Systemic Intervention: Philosophy, Methodology, and Practice*. New York: Kluwer/Plenum.
- Midgley, G. 2003a. *Systems Thinking*. Volumes I-IV. London: Sage.
- Midgley, G. 2003b. Science as systemic intervention: Some implications of systems thinking and complexity for the philosophy of science. *Systemic Practice and Action Research*, 16: 77–97.
- Midgley, G and Floyd, M. 1990. Vocational training in the use of new technologies for people with disabilities. *Behaviour and Information Technology*, 9: 409–424.
- Midgley, G Munlo, I and Brown, M. 1997. *Sharing Power: Integrating User Involvement and Multi-Agency Working to Improve Housing for Older People*. Bristol: Policy Press.
- Midgley, G, Munlo, I, and Brown, M. 1998. The theory and practice of boundary critique: Developing housing services for older people. *Journal of the Operational Research Society*, 49: 467–478.
- Mingers, J C. 1980. Towards an appropriate social theory for applied systems thinking: Critical theory and soft systems methodology. *Journal of Applied Systems Analysis*, 7: 41–50.
- Mingers, J C. 1984. Subjectivism and soft systems methodology – A critique. *Journal of Applied Systems Analysis*, 11: 85–103.
- Mingers, J and Brocklesby, J. 1996. Multimethodology: Towards a framework for critical pluralism. *Systemist*, 18: 101–131.
- Mingers, J and Gill, A (eds). 1997. *Multimethodology: The Theory and Practice of Combining Management Science Methodologies*. Chichester: Wiley.
- Miser, H J and Quade, E S (eds). 1985. *Handbook of Systems Analysis: Overview of Uses, Procedures, Applications and Practice*. New York: North Holland.
- Miser, H J and Quade, E S (eds). 1988. *Handbook of Systems Analysis: Craft Issues and Procedural*

- Choices. New York: Wiley.
- Morecroft, J D W. 1985. Rationality in the analysis of behavioral simulation models. *Management Science*, 31: 900–916.
- Morecroft, J D W and Sterman, J D (eds). 1994. *Modeling for Learning Organizations*. Portland OR: Productivity Press.
- Morgan, G. 1982. Cybernetics and organization theory: Epistemology or technique? *Human Relations*, 35: 521–537.
- M’Pherson, P K. 1974. A perspective on systems science and systems philosophy. *Futures*, 6: 219–239.
- Optner, S L (ed). 1973. *Systems Analysis*. Harmondsworth: Penguin.
- Prigogine, I 1987. Exploring complexity. *European Journal of Operational Research*, 30: 97–103.
- Quade, E S and Boucher, W I. 1968. *Systems Analysis and Policy Planning: Applications in Defence*. New York: Elsevier.
- Quade, E S, Brown, K, Levien, R, Majone, G, and Rakhmankulov, V. 1978. Systems analysis: An outline for the IIASA international series of monographs. *Journal of Applied Systems Analysis*, 5: 91–98.
- Roberts, N, Andersen, D F, Deal, R M, Grant, M S and Schaffer, WA. 1983. *Introduction to Computer Simulation: A System Dynamics Modeling Approach*. Reading MA: Addison-Wesley.
- Rosenhead, J (ed). 1989. *Rational Analysis for a Problematic World*. Chichester, Wiley.
- Scarrow, P K and White, S V. 2002. Lucian Leape and healthcare errors. *Journal for Healthcare Quality*, 24(3): 17–20.
- Senge, P. 1990. *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday/Currency.
- Shannon, C E. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27: 379–423.
- Shannon, C E and Weaver, W. 1949. *The Mathematical Theory of Communication*. Urbana IL: University of Illinois Press.
- Simon, H A. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society*, 106: 467–482.
- Sterman, J D. 1994. Learning in and about complex systems. *System Dynamics Review*, 10: 291–330.
- Taket, A and White, L. 1997. Working with heterogeneity: A pluralist strategy for evaluation. *Systems Research and Behavioral Science*, 14: 101–111.
- Taket, A and White, L. 2004. Playing with PANDA: The cyborg and the rhizome. In, *Community Operational Research: OR and Systems Thinking for Community Development*. Midgley, G and Ochoa-Arias, A E (eds). New York: Kluwer.
- Thomas, A R and Lockett, M. 1979. Marxism and systems research: Values in practical action. In, *Improving the Human Condition*. Ericson, R F (ed). Louisville KY: Society for General Systems Research.
- Trist, E L and Bamforth, K W. 1951. Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4: 3–38.
- Trist, E L, Higgin, G W, Murray, H, and Pollock, A B. 1963. *Organizational Choice*. London: Tavistock.
- Ulrich, W. 1981. A critique of pure cybernetic reason: The Chilean experience with cybernetics. *Journal of Applied Systems Analysis*, 8: 33–59
- Ulrich, W. 1983. *Critical Heuristics of Social Planning: A New Approach to Practical Philosophy*. Berne: Haupt.
- Ulrich, W. 1987. Critical heuristics of social systems design. *European Journal of Operational*

- Research, 31: 276–283.
- Ulrich, W. 1988. Churchman's "process of unfolding" – its significance for policy analysis and evaluation. *Systems Practice*, 1: 415–428.
- Ulrich, W. 1993. Some difficulties of ecological thinking, considered from a critical systems perspective: A plea for critical holism. *Systems Practice*, 6: 583–611.
- Ulrich, W. 1996. *Critical Systems Thinking for Citizens: A Research Proposal*. Centre for Systems Studies Research Memorandum #10. Hull: Centre for Systems Studies, University of Hull.
- Vennix, J A M. 1996. *Group Model Building: Facilitating Team Learning using System Dynamics*. Chichester: Wiley.
- Weaver, W. 1948. Science and complexity. *American Scientist*, 36: 536–544.
- Weil, S. 1998. Rhetorics and realities in public service organizations: Systemic practice and organizational learning as critically reflexive action research (CRAR). *Systemic Practice and Action Research*, 11: 37–62.
- Wiener, N. 1948. *Cybernetics*. Cambridge, MA: MIT Press.
- Zammuto, R F. 1982. *Assessing Organizational Effectiveness: Systems Change, Adaptation, and Strategy*. Albany NY: SUNY Press.

A Systemic Evaluation of an Agricultural Development: A Focus on the Worldview Challenge

Richard Bawden

At a time in life when the golf course looks attractive, Richard constantly shuttles between the US, Australia and India. In this chapter, Richard reveals what is behind the journey – the desire to encourage a nation's farming sector to think differently about what it does. He argues that using systems ideas is not an elite game, but something that can be played out by ordinary people doing extra-ordinary things. Mirroring Gerald Midgley's ideas, and drawing directly on Dan Stufflebeam's categorization of evaluation approaches, he describes the task in both systems and evaluation terms. On the way he lays down the following challenge to systems and evaluation practitioners as well as funders of programs. Can we expect people to think and act differently if we are not prepared to do so ourselves?

Systemic premises and assumptions

The field of endeavor of systemics incorporates systems thinking and systems practice into an intellectual and practical tradition that is frequently referred to as the systems approach. There is however, neither a single systems approach to anything – be it management, research, development, evaluation or whatever – nor an homogeneous intellectual tradition for systemics itself. What is common across all of the diversity in the field however, is an acceptance of three basic premises:

1. That any whole bounded entity (concrete or abstract, real or assumed) has properties that differ from, and that are unpredictable from a study of its inter-connected component parts.
2. That the component parts themselves are systems, and are thus regarded as sub-systems of higher order (supra) systems with their own sub-systems.
3. That the emergent properties of all systems are outcomes of the processes of the inter-connections between their sub-systems, and between the systems and the supra-systems in which they are embedded.

Accordingly, in appreciation of such an hierarchical system of systems, systemists see themselves working concurrently in three dimensions: They always concern themselves, at least conceptually, (a) with the system under review, (b) with the sub-systems which comprise that system, and (c) with the supra-system which represents the environment in which the system under review is embedded (Bawden 2005).

All systems approaches to evaluation in their practical application, reflect these basic epistemic assumptions and this three dimensional approach to development and its evaluation alike.

While a number of different typologies of systems approaches have been

promoted over the years, particularly with respect to different paradigms of social research and management science, one of the most popular and frequently cited schemas recognizes distinctions between three different camps or schools of systems methodologies: the *hard*, the *soft* and the *critical* (Jackson 2000). Each of these three schools finds expression in different approaches to development, and in the processes that are used in the evaluations of such development. The different stages can be seen to represent a development sequence of three progressive waves of systems thinking (Midgley 2000) where each new wave has emerged in response to critical evaluations of the logic and methods of its predecessors. In the first wave, the concern is for improvements in the state and performances of systems “out there” in the world. In the second wave, the focus of attention shifts to particular systems of inquiry into problematic situations in the ‘world out there’ where improvements to those situations come about through debates about systemically desirable and feasible changes. The third wave embraces the idea of systems of intervention that involve stakeholder critiques of, and actions to remedy, both the social conditions in which different categories of stakeholders find themselves embedded and the boundary and value judgments that are being made, ostensibly on their behalf, by social planners or other interventionists. The focus of improvements under these circumstances results from communicative actions that deliberately confront coercive and otherwise power-limiting constraints.

Each of the wave shifts in this sequence represents an expression of some fundamental epistemic transformation or another of basic ontological, epistemological and/or axiological positions that, while still leaving the basic systems premises intact, dictates the need for very different methodologies for each wave. Jackson (1987) proposes that each of these three different types of systems approaches is appropriate to particular circumstances, and that this indicates the utility of methodological pluralism – or horses for courses, as it were. First-wave systemic methodologies are really useful whenever there is general agreement about the nature of what it is that needs to be improved. Second-wave methodologies are relevant when the focus of improvement is much more conjectural, but where disagreements between stakeholders is typically non-coercive. And third-wave systemic methodologies find their particular use in situations which are characterized by coercion and power asymmetries. Further details of these distinctions are to be found in the previous chapter by Gerald Midgley.

The power of each one of these three waves can itself be greatly enhanced by the embrace of critical and epistemic self-reflection by those who adopt methodologies relevant to any of the three perspectives. The utility of even the most passive, non-participative, hard systems approaches to evaluation can be greatly enhanced for instance, if the systemic logic and the systems nature of all of the activities are made transparent to all involved in the project being evaluated.

Interestingly enough, there would appear to be at least *prima facie* correlation

here between these three waves of systems thinking just outlined (and the respective classes of methodologies that they each employ) on the one hand, and the three broad categories or clusters of evaluation approaches that have been identified by Stufflebeam (2001), on the other. Thus (a) the questions/methods-oriented approaches are certainly applicable to hard systemics, (b) the improvement/accountability approaches have particular relevance to soft systemics in many regards, and (c) the social agenda/advocacy approaches seem well suited to critical systemics. Furthermore, there is a strong accord between the nature of the sequence of progression of the waves of systemics and the claim by Stufflebeam (2001), that the trend in evaluation approaches is (also) toward a “strong orientation to stakeholder involvement and the use of multiple methods”.

This then brings us back to the crucial significance of worldviews among stakeholders and evaluators alike while leading to the formulation of four key assertions that draw on the observations of a number of different researchers – including Perry (1968) on intellectual and moral development, Kitchener (1983) on levels of cognitive processing, and Salner (1986) on systemic competencies:

1. That worldviews themselves can be seen to have characteristic sequences of development that involve both intellectual and moral dimensions which, however, only progress under particular circumstances of existential challenge.
2. That as acts of development in the material and social worlds are functions of the intellectual and moral development of the actors involved in them, then such epistemic development should be focus of development initiatives in circumstances where currently prevailing worldviews are limiting to further progress.
3. That the understanding and application of systems approaches to development, and therefore also to the evaluation of initiatives, demand an advanced stage of progression in the intellectual and moral development of the actors involved in both of these endeavors that is necessary to support their acquisition of systemic competencies (of thinking and practice).
4. That if intellectual and moral development of multi-stakeholder actors to achieve systemic competencies is to be a central focus of attention for development initiatives, then evaluation approaches appropriate to this aim need to be developed.

In other words, the pre-condition for adopting systemic perspectives on development and on its evaluation for stakeholders, is their epistemic development: To understand and deal systemically with complexity, one needs to have developed to a reasonably advanced epistemic state that is appropriate to dealing with complexity (Salner, 1986).

These four assertions constitute the foundational support for both the logic and the practices of systemic evaluation for epistemic development as will now be illustrated by a case-study from Southern India.

Evaluating epistemic development – A systemic approach

A project that is being conducted by academics from Michigan State University (MSU) and Tamil Nadu Agricultural University (TNAU) in collaboration with a wide spectrum of other stakeholders in horticultural development from producers to consumers is a truly systemic endeavor in a number of different ways. Essentially it is focused on the transformation of worldviews from productionism to a broader systemic perspective across broad constituencies of stakeholders of a number of various horticultural food systems. The central assumption is that such a transformation is a function of the intellectual and moral development of those the stakeholders, which thus becomes the aim of the initiative and the focus of its evaluation.

The wide variety of approaches that are available to evaluators of any development initiative today is an expression of the influence of a combination of differing theoretical assumptions, philosophical premises, practical objectives, and contextual appraisals. The set of profound beliefs that each evaluator holds as his or her worldview about the nature of nature (ontology), the nature of knowledge (epistemology), and the nature of human nature (axiology), is reflected in the approaches that he or she chooses to employ in practice – knowingly or unknowingly, consciously or unconsciously. Given the paramount influence that the worldview perspective that any individual evaluator brings to bear in any particular exercise of evaluation, it is not only regrettable when the issue of perspectives remains unaddressed, but also grossly negligent. If indeed, as Stufflebeam (2001) argues, any evaluation is a *study that is designed and conducted to assist some audience to assess an object's merit and worth*, then explicit attention must be paid to foundational assumptions about the nature of *worth* and *value*, and to how these can come to be known in any given contextual situation, if it is to be an ethically defensible practice.

This is most especially so when the objective of the program or project to be evaluated is itself focused on the transformation of worldviews among and across a multi-stakeholder constituency. Such is the circumstance that increasingly faces evaluators of initiatives within so called international development, where the fundamental aims and objectives of programs and projects – and indeed the very ethos of the enterprise – are slowly undergoing a radical sea-change to reflect what Sen (1999) refers to as *development as freedom*. From this rights-based perspective, Sen insists, the success of any society is to be evaluated by the extent to which the citizens within that society are *capable of living the lives that they have reason to value*; and this in turn highlights the primary significance of evaluative consciousness and active appreciation of *merit* and *worth* as central aspects of the development process. This is about as far removed from the productionist view of development as modernization as one can possibly imagine and indeed represents nothing less than a revolution in prevailing worldview perspectives: such a revolution must extend beyond those agents and agencies that assume some responsibility for

nurturing such a shift in perspectives and for evaluating associated initiatives, but must also include the very broad spectrum of other stakeholders who are involved in agri-food systems, as they are presented with opportunities to recast and re-evaluate their own perspectives on what constitutes better.

Important insights into the way by which post-productionist programs can be designed, conducted and evaluated, are provided by some ideas from the field of systemics – with its focus on wholeness, connectedness, embeddedness, and emergence. This claim is illustrated in the particular multi-methodological systems approach to evaluation that is described below, and which might be summed up as “an approach to assess the merit and worth of the development of competencies in systems thinking and practices by multiple stakeholders who are involved across whole agri-food systems”.

The project, in which the author is an active participant/evaluator, presents a powerful example of the inseparability of development and evaluation – with self-evaluation of personal and collective epistemic development being an integral aspect of the development process. The project has been designed explicitly using systems principles that express aspects from all three of the schools of systems thinking. The claim of the project is that through a range of activities including interviews, workshops, network initiatives, commercial relationship development, and other activities, stakeholders will become more systemically competent as expressed in their approach to the further systemic development of particular horti-food systems. In the process, stakeholders will also become explicitly conscious of their own epistemic development through their critically reflexive involvement in project activities. The aim of the evaluation then is to assess the *merit* and *worth* of these claims as assessed essentially by the stakeholders themselves through a process that itself further contributes to the development of those systemic competencies and to epistemic awareness and development.

THE CONTEXT

Over the past four decades or so, India has made quite remarkable strides in achieving national food security. From a situation where it regularly had to import grain to provide the staple food needs of its citizens, the country is now able to meet the needs of a greatly expanded population from those times, through its own production: Indeed the national grain stocks currently have many months-worth of home grown grain in reserve.

A major key to the outstanding success of this *green revolution* has been the widespread adoption by farmers of those science-based technologies and economic management techniques that represent the twin pillars of industrial agriculture. And central to this process of agricultural development, has been the role of Indian institutions of research and of higher education not only in extending technologies and management techniques to farmers, but also in promoting the ways of thinking that reflect what might be referred to as the productionist worldview. Governance institutions have also played key roles in

this regard, by formulating policies and regulatory strategies that, in privileging commodity production as the primary aim of agricultural development, have further enhanced the merits of the productionist worldview and in essence, rewarded its adoption.

Every particular worldview reflects a specific set of what can be referred to as epistemic assumptions: assumptions about the nature of reality, the nature of knowledge and the nature of human nature as it relates to matters of values, power, relationships, and so on. To adopt the epistemic assumptions and beliefs of any one worldview is not only to generally exclude other perspectives, but to also seriously impede any attempts to seek further alternatives – even when the need for these is indicated by circumstances.

While the widespread adoption of productionist worldview mindsets or mental perspectives has been a vital aspect of the huge amplification of agricultural production in India, this has not been without some considerable costs. The epistemic foundations of productionism lie in the reductionism and objectivism of techno-science and of neo-liberal economics. The most obvious restriction of such foundations has been the difficulty of accommodating broader, more systemic concerns for those negative environmental and socio-cultural impacts that have so often accompanied the intensification of agricultural production. Thus the very pervasion of productionism in India has resulted in significant epistemic impediments to the adoption of more systemic or holistic worldviews of development that are more appropriate for dealing with the complexity of the situation in Indian agriculture.

THE CHALLENGE

One of the most significant challenges facing Indian institutions is therefore the design, conduct and evaluation of participative development strategies that lead to the adoption of post-productionist worldviews (without imperiling food production) and to the nurturing of systemic appreciation and holistic worldviews among the citizenry, and within social institutions, alike. This challenge has given rise to the search for more systemic approaches to agricultural development that reflect much broader perspectives than those promoted by productionism, and to processes that facilitate the transformation of the prevailing worldview through exposure to different epistemic assumptions.

The fundamental focus of development is thus shifted from acts of development in the material and social worlds of citizens to the epistemic development of those citizens. And this is a much more comprehensive and sophisticated demand than conventional technical and/or managerial education or training presents.

This position then dictates approaches to monitoring and evaluation of program and project development activities that are appropriate to the systemic nature of the claims. The particular process of systemic evaluation that is described below has its genesis in such a response, while reflecting attempts to integrate

ideas and practices from different approaches to the process of *development* itself, to *evaluation*, and to that domain of thinking and practices known as *systemics*.

The essential systemic evaluative process is a dialogic, participative, reflective, and democratic form of collective or social experiential learning, in which the wholeness, connectedness, embeddedness, and three-dimensionality of systemics is explicitly exploited in a number of different and vital ways. Thus, if the project is being successful:

1. Stakeholders should be able to express (and evaluate) how they have come to appreciate themselves collectively as a learning collective or sub-system of the developing horti-food system in which they are embedded, and which is itself embedded in turn, in higher-order environmental suprasystems with both bio-physical and socio-cultural dimensions, which presents both challenges and opportunities for the further development of the horti-food systems within systemic and sustainable contexts.
2. Stakeholders should also be able to express (and evaluate) their appreciation of the systemic nature of their own learning sub-system with respect to three dimensions of learning or cognitive processing:
 - a. They can process everyday matters-to-hand in their search to improve circumstances that seem problematic to them in the 'real world' about them (cognition).
 - b. They can process the way by which they process those matters-to-hand, in seeking to improve the way they go about our processing (meta-cognition), and
 - c. They can process the way that they frame the way they go about their processing and process of processing, in seeking to identify the worldviews that shape their framing as a prelude to transforming them if they prove to be constraints to the other two levels of processing (epistemic-cognition).
3. Stakeholders should be able to express (and evaluate) the systemic nature of the learning process that is relevant at each of these three levels – that is the manner by which the different learning activities (divergence and convergence for example) and learning modes (empirical and ethical for instance) interact with and inform each other.

When written in such formal terms, these matters sound impossibly idealistic in the context of uneducated peasant farmers with their practical and sometime mystical worldviews on the one hand, and the academic experts with their conventional techno-scientific worldviews on the other. The challenges to the systemic development agents and to the systemic evaluators alike are to frame their strategies in a manner that is sensitive to different ways of knowing and valuing, and to find ways of interpreting the conceptual ideas into appropriate everyday language.

THE RESPONSE

A natural progression of development (and thus foci for evaluation) is emerging in the Indian project that can be interpreted as an echo of the idea of the sequential development of the three waves of systemics.

1). In the first stage of the project, the development focus was on the creation of an awareness of the basic systemic nature of the horticultural industry and then of the characteristics and organization of various horti-food systems. The key development strategy here was multi-stakeholder workshops organized by faculty from TNAU and facilitated by the MSU team, in which the participants worked in groups to evaluate, from their own particular perspectives, the strengths (S), weaknesses (W), opportunities (O) and challenges (C) across the entire value chain of specific horticultural commodities that were of current relevance to them. Aspects of the resource base were also considered in this way as were the socio-political climates that prevailed. The participants were encouraged to discuss and debate within their groups, any individual differences in their situational assessments, and to reflect on the possible reasons behind those differences. The full profiles were then explored in plenary sessions.

In addition to gaining vital information about the state of the system from the participants, the key aim of this first stage of the development project was to deliberately create an awareness of the inter-connected systemic nature of the horticultural industry: of how the various component parts of the value chains of horti-food systems interact with each other, with the system as a whole and with the environments in which it operates. Participants were also exposed to the idea that the horti-food system is a conceptual construct, and can be analyzed for various qualitative characters, in a manner (using the SWOC criteria) that invariably illustrates different perceptions of the same reality.

In essence the particular systemic focus of this first phase reflected first wave systems thinking of *systems out there*, and was thus evaluated in those terms: With appropriate question/response activities, it was not difficult to determine the extent to which participants came to appreciate the complex of relationships that were associated with different elements of the value chain, with the resource base, with the policy and regulatory climates, and with the bio-physical and socio-cultural aspects of the environment. The evaluation criteria used here reflected the general features of Stufflebeam's Questions and Methods Oriented Evaluation Approach with the assumption that 'it is usually better to answer a few pointed questions well than to attempt a broad assessment of a program's merit and worth' (Stufflebeam 2001). While the questions focused on participant appreciations of the systems nature of the horticultural industries with which they were most familiar, the method explored was essentially the SWOC process and its outcomes.

2). As the project progressed into its second phase, the focus was essentially on the generation of suggested strategic activities perceived, by the participants, as improvements to the current situations within the horti-food systems that have been explored. The particular emphasis at this stage was on the significance of the

relationships between the judgments of betterment by different stakeholders and the worldviews that they held. What strategic action might represent an improvement for one person (or for one entire category of stakeholders) might actually be a dis-improvement for others: what is a better state for some, can be a worse state for others: And this type of perception analysis is given further focus when the possible socio-cultural and ecological impacts of potential strategic changes are explored by the participants. Issues to do with differing perceptions within multi-stakeholder groups were thus further reinforced at this stage of the development project, particularly as they related to the vital interconnections between worldviews and what were considered to be desirable and feasible transformations.

The systems emphasis in this stage therefore, was more on the significance of inter-relationships between participants and their ideas, than on the horti-food systems *per se*. This is akin to the shift in systemicity “from systems of the world to systems for inquiring into the world”, of which Checkland (1981) talks: or the change from the wave of hard systems thinking to the soft systems thinking wave.

A vital element of the development project now becomes the growth of consciousness of the systemic nature of the processes of judgments, of learning, of development, and of evaluation, in the sense of the integrated notion of a whole process and of the dynamic inter-connectedness of the various parts that constitute that process. A modified soft systems methodology (SSM)¹ proves useful for both the project activities and for evaluation purposes at this juncture, with its emphasis on the inter-relationships between the various activities that together constitute the whole process, and its specific attention to models of human activity systems abstract constructs, and not concrete entities. Evaluations of this stage of the process, with its distinct shift from the concrete to the abstract and the objective to the normative, are best conducted by replicating the SSM as a vehicle for establishing what different stakeholders feel about the nature of improvements to their world through development activities.

In this manner, this stage of systemic evaluation is consistent with Stufflebeam’s claim for Improvement/Accountability Oriented evaluation approaches which emphasize the need for “comprehensiveness in considering the full range of questions and criteria needed to assess a program’s value” (Stufflebeam 2001). It is also congruent with the need, with this mode of evaluation, to “engage stakeholders in focusing the evaluation” and with the grand claim of “subscription to the principles of a well-functioning democratic society” which, by its very nature not only tolerates differences but exploits their inherent synergy. On the other hand, there is a stark contrast between this category of evaluation approaches, and the systemic thrust being promoted here with Stufflebeam’s insistence that the philosophical foundations of the Improvement/Accountability Oriented approaches to evaluation “include an objectivist orientation to finding best answers to context-limited questions”. Soft systemics is

1 See the Chapters by Boon Hou Tay, Ken Meter, and Kate Attenborough.

anything but objectivist in its foundations, and is orientated towards the search for “betterment” and never “the best”.

The crucial point to be emphasized here is that the systemic evaluation at this phase of the project, is conducted in a participative and discursive style, and is regarded as an integral aspect of the development project itself. In this context, it makes sense to think of all the actors involved in the process – development agents, stakeholders, and evaluator – as a coalition of self-evaluators, or better yet, as the reflective function of what might now be seen as a learning sub-system of horti-food systems! The evaluative questions that they must ask of themselves, with the assistance of the evaluator, address not just questions about the merit and worth of the raising of systemic awareness among them, but also of the merit and worth of the actual judgments that they are making with respect to improvements to their horti-food systems.

3). The third stage of the project, which is only now being introduced, takes the systemic connection between worldviews and specific opinions and ideas about improvements to horti-food systems, to a more advanced level: Now the focus is on what we can refer to as epistemic development. This builds on the arguments rehearsed earlier that (a) our epistemic status – the particular nature of the worldviews that we assume – is itself capable of development, and (b) that we need to develop quite advanced epistemic states to really grasp the significance of what it means to approach complex issues of development from a profoundly systemic perspective. The aim of this third phase of the project in India therefore, is to help stakeholders become increasingly aware of the strengths, weaknesses, opportunities and challenges of different epistemic positions as foundations for effective debates about desirable and feasible changes, in contrast to the much more limited and superficially systemic focus on strengths, weaknesses, opportunities and challenges of the development of horti-food systems.

This epistemic investigation poses considerably higher levels of challenge to the evaluating coalition of stakeholders – with the clear implication from theory and empirical observations alike, that until and unless there has been critical epistemic development within that coalition, even the concept of such development will often be exceptionally difficult to grasp. Similarly, until and unless such epistemic development reaches a particular level of maturity, the systemic idea will invariably remain elusive. This third phase of the project is set clearly within a critical context from three key perspectives:

1. Current socio-economic conditions within the rural areas are such that many citizens are disempowered from seeking to live the lives that they have reason to value – this development project is therefore emancipatory in its character.
2. Conventional approaches to development in this area reflect systems of human activities that are bounded by decisions made by bureaucrats and technical experts within the context of productionist worldviews – this development project is therefore concerned with explicit critiques and

transformations of such boundary judgments.

3. The prevailing productionist worldviews impede the development and adoption of more comprehensive systemic perspectives on improvements to the existing conditions – this development project seeks to nurture the epistemic development of systemic worldviews across the spectrum of stakeholders involved.

There is certainly some in-principle congruence here with the third category of Stufflebeam's typology of evaluation approaches – Social Agenda/Advocacy – in the claim that he makes for them as having 'an affirmative action bent' in which program evaluation is used "to empower the disenfranchised". There is also a strong correspondence between this class of approaches to evaluation and critical perspectives of development, with the assertion that "they are strongly oriented to democratic principles of equity and fairness and employ practical procedures for involving the full range of stakeholders". Evaluative questions, generated through reflexive discourse with stakeholders, and focused on the democracy of representation and on the quality, authenticity, equitability and extent of the deliberation, can certainly capture the essence of the first two of the critical perspectives above. They do little to address the third, most vital aspect mentioned above however: Nothing that Stufflebeam presents within his category of Social Agenda/Advocacy approaches to evaluation allows judgments to be made about the merit and value of worldview transformation – nor are there any hints as to how such a "deliverable" might be monitored and evaluated.

In the absence of detailed conceptual and methodological guidance from existing approaches to evaluation therefore, this last part of the project in India, is very much work in progress. And at its heart lies a difficult conundrum: For our team of evaluators to assist in the development and judgment of criteria related to the transformation of worldviews to accommodate profoundly systemic perspectives on the world – essentially the facilitation of stakeholder development as "systemic beings" – we ourselves need to undergo such an epistemic transformation as a precondition.

The logic presented here, I believe, dictates that such a competency is imperative in the face of the complex challenges of epistemic transformation for systemic development.

References

- Bawden, R.J. 2005. Systemic Perspectives on Community Development: Participation, learning and the essence of wholeness. *Perspectives on Community Development in Ireland* 1: 45–62.
- Checkland, P.B. 1981. *Systems Thinking Systems Practice*. John Wiley and Sons. Chichester.
- Jackson, M.C. 2000. *Systems Approaches to Management*. New York: Kluwer Academic/Plenum Publishers.
- Jackson, M.C. 1987. New Directions in Management Science. In, *New Directions in Management Science*. M.C. Jackson and P. Keys (eds). Gower: Aldershot.

- Kitchener, K S. 1983. Cognition, metacognition, and epistemic cognition: A three level model of cognitive processing. *Human Development* 26: 222–232.
- Midgley, G. 2000. *Systemic Intervention: Philosophy, Methodology, and Practice*. New York: Kluwer Academic/Plenum Publishers.
- Perry, W G. 1968. *Forms of Intellectual and Ethical Development in the College Years*. New York: Holt, Rinehart and Winston.
- Salner, M. 1986. Adult cognitive and epistemological development in systems education. *Systems Research* 3: 225–232.
- Sen, A. 1999. *Development as Freedom*. New York: Anchor Books.
- Stufflebeam, D L. 2001. Evaluation Models. *New Directions for Evaluation* Issue 89, Spring 2001: 7–105. San Francisco CA: Jossey-Bass.
- Yankelovich, D. 1991. *Coming to Public Judgment: Making democracy work in a complex world*. Syracuse NY: Syracuse University Press.

System Dynamics-based Computer Simulations and Evaluation

Daniel D Burke

It is easy to categorize systems dynamics as program logic with bent arrows. However, that barely touches on the power of systems dynamics to reveal insights into programs, especially in situations where the situation is subject to multiple and often delayed feedback. What usually emerges from these situations is often dismissed by program administrators and evaluators alike as “unintended” or “unforeseen” consequences. This marginalizes substantial learning and valuing opportunities. Dan’s example of evaluating a teachers’ professional development program is superficially a mainstream evaluation task, but he reveals insights that most evaluation methods would barely touch. Although the focus of his example is at the program planning stage, system dynamics can and has been used as analytical tool for answering questions such as “why did the program behave in that way”, “how valuable was this part of an intervention relative that that part of the intervention”.

The AEA Policy and Procedure Manual defines evaluation as follows: “**Evaluation** involves assessing the strengths and weaknesses of programs, policy, personnel, products and organizations to improve their effectiveness.” www.eval.org/aea100103.pp.pdf. Using this definition, I will discuss the use of a program evaluation tool, system dynamics-based computer simulations (SDCS), for understanding the manner in which the interaction of a program’s structural components generate the program’s behavior over time and the relationship between its behavior and whether the program achieves its goals. SDCSs are appropriate when the problem situations are aggregate, interrelated, and continuous. In such instances, computer models can have the advantage of making the assumptions concerning the program explicit, logically computing the consequences of the assumptions, and are able to interrelate the behavior of multiple factors simultaneously in a way the people cannot. (3) As a program is being developed, the use of SDCS enables one to assess the possible consequences of differing policy options and/or resource allocations against defined program goals, thus contributing to the improvement of the program design. Unintended consequences of the policies/allocations may be revealed as well.

When the program is functioning and after it has been completed, SDCS enables an evaluator to assess its strengths and weaknesses and how these impact the extent to which the program achieved its goal(s). SDCS is effective if the program goals are quantitative or qualitative. An example of a quantitative goal would be changing the teaching behavior of a specific number or percentage of teachers who have engaged in a particular professional development program. A qualitative example would be when the goal is achieving the best possible outcome of a particular policy and resource environment – such as identifying the particular

allocation of available resources and set of policies to enable a school district to obtain the most effective teaching workforce.

SDCS, then, can be a powerful tool enabling us to evaluate and refine program design at the program's inception, to assess and increase the effectiveness of a program as it is implemented, and, when it is completed, to learn why and to what degree a program succeeded or failed – and make judgments of worth on those bases. In the SDCS described in this paper, questions such as, “should a workshop be required of all teachers or open to volunteers” and “what is the likely consequence of providing different numbers of workshops and thus training different numbers of teachers” were asked. The answers to these questions allow districts to design more effective professional development programs that make better use of their resources.

Problems Evaluating Multicomponent Programs

Frequently, in identifying and evaluating the strengths and weaknesses of critical factors in a multicomponent program, we use a jigsaw-puzzle graphic to illustrate that the factors interact. This representation is similar to the matrices often used to score implementation of program or system components. While there may be an overall program goal, each program factor typically is represented by a series of steps/levels that allow us to evaluate, or benchmark, the degree of implementation of each individual factor. However, neither the graphic organizer nor the matrix enables us to understand in what ways the factors interact or to integrate the functioning of the separate components. If there are eight components with five levels of implementation, does this mean that being at the top level in the first four components and the bottom level in the last four components will produce the same program outcomes as a reciprocal score, or the same outcomes as being at level three in all components?

Several simple analogies may highlight the difference in component versus program function. Depending upon their relationship, if each of eight components of a program is functioning at 90% efficiency, the program itself may be functioning at less than 30% efficiency. To illustrate this, consider that when stacking blocks you only need be off by a little in placing each block on top of the next for the pile to collapse. Even with an overall measure of program function, these graphical representations or matrix measures do not enable one to understand the nature of the mismatch between program component function and overall program function. Also, this approach to measuring program or system function produces a static snapshot that does not illustrate the program or system's likely trajectory. That is, is the program effectiveness likely to increase? Decrease? Will unforeseen outcomes arise? What will be the likely result if a policy or resource change is made?

THE SYSTEM DYNAMICS APPROACH

The system dynamics approach to social systems analysis grew out of nonlinear dynamics and feedback control theory. (1, 2, 4) Using a system dynamics approach entails attempting to identify the interactions between the system's key

components, often called its stocks, and the flow between stocks and asks whether these interactions generate the behavior being exhibited by the system. The system dynamics approach assumes that these components interact through feedback loops and that delays often occur in either material or information flows through the system. The approach is particularly concerned with the behavior of the system over time.

As the examples below illustrate, feedback loops refer to a situation of X affecting Y and Y in turn affecting X, perhaps through a chain of causes and effects. Thus, we cannot study the link between X and Y independently of the link between Y and X to predict how the system will behave. Only by studying the feedback system as a whole will we be able to develop meaningful understanding.

Feedback loops may be either reinforcing (positive) or balancing (negative). Population growth is an example of a reinforcing feedback loop, the larger the population the greater the increase in population leading to an even larger population and an even greater increase. A positive loop for population can also function in the opposite direction resulting in a decrease in population. Currently, several countries have a birthrate of less than two per family. This will result in an even smaller population. Certainly a population control program needs to account for these separate types of positive feedback loops.

A thermostat is an example of a balancing feedback loop. In the winter as the temperature gets lower than the thermostat set point the furnace is turned on until the temperature reaches the set point at which time it is turned off. Then as the temperature drops below the set point the furnace is again turned on. This example also illustrates the concept of goal setting. A particular temperature is set as the heating/cooling system goal and it is the gap between this set point and the actual temperature that drives the system. In a variety of social programs and organizations, one may see such goal-based balancing loops. A program goal is set, such as some desired level of product quality or a certain percentage of students passing a test (as is now required by the “No Child Left Behind Act”). If the goal is not met, the gap between the actual performance and the goal is used as an argument to obtain more resources to close the gap. As the gap closes, fewer resources are added. If it opens again, more resources are added.

A second important concept in system dynamics analysis is that of delay, either in information or in material flows through the system. Such delay is the cause of oscillatory behavior in the system. A typical instance of predator/prey behavior illustrates this. Wolves don't know when the rabbit population is going down and continue to have pups, thus building the wolf population. This results in an overshoot of the wolf population – too many wolves – and not enough rabbits, resulting in decline in the wolf population from starvation because of the smaller food supply. With a decrease in the wolf population, the rabbit population can rebuild triggering the cycle again.

Oscillations due to delay in information flow often occur in labor markets. A particular profession is thought to have a lot of available jobs. This entices

people into the profession and its population grows. However, information concerning the remaining need in the profession is slow in reaching the labor market. People continue to go into it and only to find there are no jobs. The number of entrants then drops sharply so that more people leave the profession than enter it, resulting in a new wave of job openings. Eventually the information emerges that there is a gap and the cycle enters another oscillation. This has certainly been seen in the job market for teachers and should be taken into account in the implementation of teacher preparation programs.

In the SD perspective, the behavior of the system over time is generated by the interaction of these feedback loops and delays. Most important from an evaluation perspective, this system behavior is often nonlinear, resistant to change, counterintuitive and/or has unintended consequences on the system outcomes.

The system dynamic approach is of greatest value in complex programs. A computer simulation (SDS) is necessary to make the most effective use of SD because we simply cannot predict the likely outcomes of nonlinear interaction of multiple system components over time. I believe SDCS is a particularly useful evaluation tool in the sense that it allows us to evaluate a program's design prior to its start. SDCS's generate scenario of how the various assumptions being made about system components will interact to produce the general behavior of the system and how this system behavior may change over time. Further, we can relatively easily test the impact of various policy and resource allocation decisions on the program outcomes using SDCS, and thus optimize the program's design when actually put into practice. After the program is implemented – a situation more commonly faced by evaluators – we can continue to use SDCS to help understand and evaluate why the program is or is not behaving as planned and evaluate the likely effect of policy or resource allocation changes will be on the system's behavior.

A particularly important use of SDCS is to examine the unintended consequences that a policy may generate and thus evaluate the decisions made by program participants. A recent example of unintended consequences in education is the outcome of mandating small class size in the K12 school districts in California. Mandating smaller classes required increases in the number of classes, thus requiring more teachers in each system. Since the production of teachers by traditional preservice programs requires at least four years of training, there were an insufficient number of teachers available to meet the needs of all schools after this policy change. Suburban districts with greater resources met the increased demand by recruiting experienced teachers from urban districts. Thus, not only did the large urban districts lose many of their best teachers, they could only fill their need for more teachers with uncertified teachers resulting in a less-qualified workforce than before the mandate. Since more classes require more classrooms and there is a large delay in producing new classrooms, a second unintended consequence of the mandate was that children who had just gotten out of trailer classrooms wound up in them again.

HOW DOES ONE GO ABOUT DEVELOPING A SDCS MODEL?

My intent here is to provide the reader with a general overview of the process to help in deciding how much more they would like to know. If a reader wishes to follow up and learn more about SDCS the literature cited and resources at the end of this paper are a good start and should be checked.

You will need software to develop and build an SDCS. There are several easy-to-learn software programs available, including Vensim, Stella, and others. These are basically drag-and-drop graphical programs that don't require learning a programming language or writing complex equations to build the model. Generally the software providers offer either free limited-function or demonstration packages available for download. Obtaining demonstration copies of these packages and trying them out is a good way to decide whether you wish to pursue SDCS building as a tool in your work.

While the software is easy to use, be forewarned that, as with any methodology, it will take a considerable amount of study and much practice to become adept at developing effective SDCS models. However, the effort invested in using some of the basic concepts of system dynamics will yield valuable insights that can guide the design and analysis of a wide variety of evaluations, especially if the situation has multiple sources of feedback, each subject to varying delays. The act of working with program practitioners in developing the causal loop diagrams described below or in working directly on a computer model is an extremely valuable tool in designing evaluations. It is similar to developing logic models but, critically, directly includes the impact of feedback loops and delays.

In using SDCS as an evaluation tool I often approach it in this way:

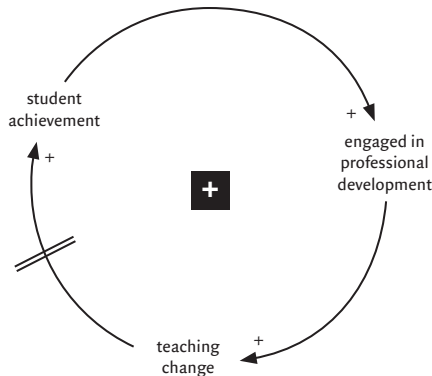
1. Identify the outcomes to be measured.
2. Use a Causal-Loop Diagram(s) (CLD) (see page 52) to develop a hypothesis or logic flow that helps to understand the program logic and how the outcomes may behave.
3. Based on the CLD, build an SDCS model of the program.
4. Test the model using the initial program assumptions and policy alternatives and refine the program design. One is using the model to test whether the program will behave as hoped when it is implemented. If it doesn't appear so to the developers, they may wish to modify the program.
5. As the program is implemented, compare the actual outcomes to those predicted.
6. Refine the model to reproduce the programs actual behavior. While in Step 4 one may attempt to predict the interaction of the system components, in this step one may better identify the component interactions that produce the program strengths and weaknesses from its actual behavior
7. Devise and test alternative policies to determine whether they may increase the program's effectiveness.

While the use of SDCS has many strengths one needs be aware of its limitations in order to make best use of the technique. It is most appropriate to use when one is interested in the behavior of the system over time, but does not yield exact point outcomes. In the example I give below, the take-home message has to do with whether the professional development mechanism being examined is likely to be effective or not. One would not go to the bank with the exact percentage of teachers being converted the simulation produced. Secondly, in developing the simulation it is critical that it be a joint effort of the evaluator(s) and the program practitioners and participants. This is necessary to develop a meaningful understanding of the model's boundaries (what to include, what to leave out), the decision rules that the actors actually follow, and the specification of the "soft" or qualitative/descriptive variables. (3) Without this team effort, one is left open to the charge that the simulation results came out the way desired by the evaluator.

CAUSAL-LOOP DIAGRAMS (CLD)

I have found that a good place to begin the development of an SDCS for evaluation of a program's behavior is with a CLD¹. A CLD is a graphic model of some of the key system variables connected by arrows that denote the causal influences among the variables. Each arrowhead is identified as either positive (+) or negative (-) to indicate how the dependent variable changes when the independent variable changes. It is important to understand that the + at the arrowhead does not necessarily mean that the dependent variable is increasing, only that it is changing in the same direction as the independent variable². Likewise, a minus symbol at the arrowhead does not mean that the dependent variable is decreasing, only that it is changing in the opposite direction from that of the independent variable. As we

Figure 1: An example of a positive (reinforcing) feedback CLD



- 1 See also the chapters by Jay Forrester, Ken Meter, and Richard Hummelbrunner for discussions about the use of CLDs.
- 2 Editors' note: This is one of the most common confusions in the systems dynamics field and important to understand.

move around the loop, what was the dependent variable becomes the independent variable. In examining the loop, if the number of negative links is even, it is a positive feedback loop as indicated by the large + sign in the center. If the number is uneven, it is a negative feedback loop as indicated by the large - sign in the center.

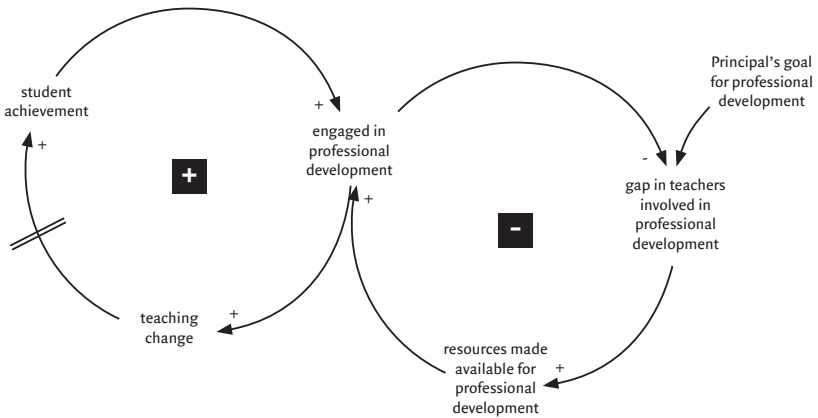
Figure 1 is an example of a positive (reinforcing) feedback CLD in which we include some of the important factors in determining what will occur as teachers engage in a professional development program that leads to a change in their teaching behavior and an increase student achievement.

The professional development program is represented at the right-hand side of the loop. The + sign at the arrowhead leading to teaching change indicates that the more professional development a teacher receives, the more her/his teaching changes. The increase in teaching change, after a delay indicated by the double hash marks as the teacher becomes more effective in using what s/he learned, leads to an increase in student achievement. The cycle continues as an increase in student achievement leads to an increase in engagement in professional development. As there are an even number of negative links ³ this CLD is an example of a positive feedback loop as indicated by the + in the center of the loop.

In Figure 2, this CLD is expanded (Figure 2) adding a negative (balancing) feedback loop that incorporates the factors that drive the implementation of the professional development program in the first case.

Here we assume that a principal (or other administrator) believes that the content of a particular professional development program is valuable and all teachers should participate in the program. This then is the principal's goal seen in the

Figure 2: CLD from Figure 1 with added negative (balancing) feedback loop



3 For this purpose we can consider zero to be an even number

upper right hand corner of the CLD. The difference between this goal and the number of teachers who have been engaged defines the professional development gap. Resources are made available to engage teachers in professional development as a function of the size of the gap. As more teachers are engaged the gap becomes smaller. As the gap decreases, the amount of resources made available decreases changing in the same direction as the gap.

This CLD could provide the basis of an SDCS to examine whether the resources made available are sufficient to engage a large enough number of teachers in professional development to lead to a measurable increase in student achievement. The SDCS would also examine the interplay between the delay in increased student achievement and the resources made available on whether the program succeeds. The logic is that if only a small number of teachers begin the program and/or if there is a significant delay between teaching change and rise in student achievement, the principal may come to feel the program is not valuable and not provide resources or the teachers may feel the it is not valuable and choose not to engage in it.

SYSTEM DYNAMICS-BASED COMPUTER SIMULATION

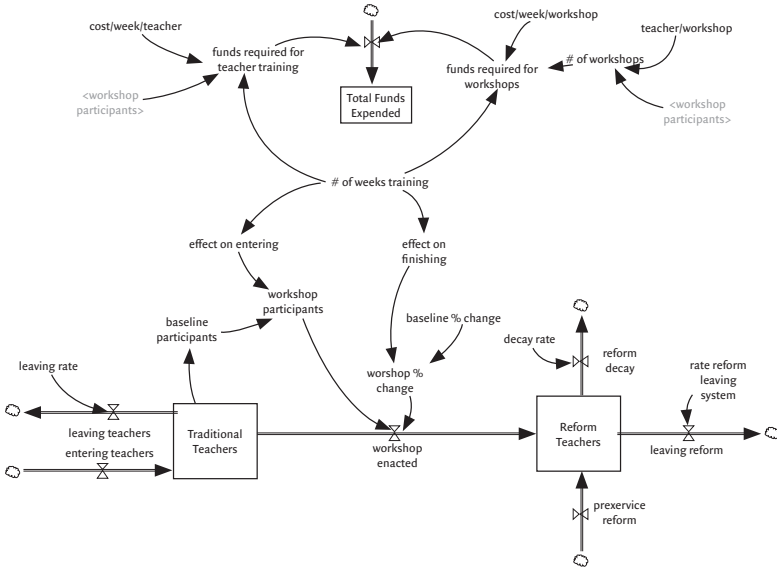
However, useful as these diagrams are, they only give a partial and possibly misleading picture of a system's behavior. Whilst they explain relationships between system components, they importantly do not show how those relationships play out over time. For that you need to simulate.

On the following page is an example illustrating an SDCS that was produced to determine whether a planned professional developed program could succeed in reaching the goal desired by the district. The SDCS also helped determine the most effective way to implement the program and tracked program costs. I have used this particular model because it illustrates the use of several non-linear variables demonstrating one of the major strengths of this modeling technique, the ability to use "soft" variables, and illustrates how even a model that contains just a few feedback loops or delays can be very useful. The example also deals with a specific, practical question that arose in many school districts that were engaged in reform of their K12 education. After discussing the development and function of the model, I will illustrate how this model can be expanded making use of multiple feedback loops and delays to make it even more powerful.

A number of school districts with which I dealt responded to the call for education reform by attempting to implement a standards-based curriculum in all of their schools and offering professional development workshops to increase teachers' ability to communicate this new curriculum to their students. The workshops were often formatted as multi-week, summer programs. However, these were not producing the hoped-for transformation of the teaching workforce. We know that teachers can learn to teach the new curriculum, so the question for a district was not whether summer workshops could build teacher capacity, but could they do so for a critical mass of teachers in a reasonable time period? If so, what factors impact this transformation? What are the costs associated with this form of professional

development? These questions are amenable to SDCS modeling since quantitative values are available for most of the important variables in the simulation and, as I will discuss, reasonable estimates can be derived for the qualitative variables. Figure 3 is the SDCS that was developed to answer these questions.

Figure 3: An illustration of an SDCS (System Dynamics-based Computer Simulation)



As you can see, this model is substantially different from the CLD. CLDs tend not to contain stocks and flows – both critical components to comprehending the dynamics of the situation.⁴ For this reason many systems dynamics modelers skip the CLD stage and go straight to SDCS. To understand this model several design conventions need to be defined. First, the boxes represent stocks or things that can accumulate. These may be objects – in this case, teachers (*Traditional Teachers and Reform Teachers*) and dollars – but they can also represent intangibles, such as knowledge or rage that can also accumulate. The double-lined arrows represent flow into or out of the stocks. In those cases where the arrows originate in or end in a cloud, the clouds are considered exogenous to the model, that is, they influence other variables in the model but are not calculated by the model (3). Here such clouds represent teachers coming into the system from any source and teachers leaving the system for some other job or system. Finally, the single line

4 See Jay Forrester's chapter for a detailed discussion on stocks.

arrows represent variables such as *effect on entering*, or constants, such as *cost/week/teacher*, that influence the flows rates.

In building this model:

1. I first established the teacher stocks, *Traditional Teaching* and *Reform Teaching*, and the flows into, between, and out these stocks. The two stocks represent the teachers who had not taken the workshop and those who had taken it and now use reform-teaching methods. The various flows not only track the movement from traditional to reform teaching, but they also track the movement of both groups of teachers into and out of the system. The latter is often not considered when a program is implemented and can be key to its success or failure.
2. I then identified and assigned values to the system constants and variables that control the flow from traditional to reform teaching. This demonstrates several important concepts illustrating the utility of this type of modeling. It is well known that when offered professional development programs, length is a critical factor, the longer the program the less likely it is that a teacher will sign up. However, it is also known that the longer they are, within limits, the more effective they are in catalyzing change in teacher behavior. Neither of these effects can be captured by a linear equation, but they can be fairly well approximated by graphs or tables. Graphs representing these variables are built and incorporated into the SDCS model as *effect on entering* and *effect on finishing*. The model is designed so that a target number of participants, *baseline participants*, is set as a constant, and the percent of workshop participants who change behavior is also set as a constant, *workshop enacted*. The length of the workshop, *# of weeks of training*, can be varied and the interaction between this length and the *effect on entering* variable determines impacts on the workshop participants, while the interaction between this length and the *effect on finishing* variable affects the effectiveness of the workshop in terms of the percentage of participants whose behavior changes. Thus, the flow-rate workshop *Traditional Teaching* to *Reform Teaching* is the product of *workshop participants* and *workshop % change*. Changing *# of weeks of training* changes the value read-out of *effect on entering* and *effect on finishing*, thus changing the value of workshop participants and of *workshop % change* and, thus, the flow of teachers from *Traditional Teaching* to *Reform Teaching*.

The question arises as to where the values for these variables (*target participant*, *baseline % change*, *effect on entering*, and *effect on finishing*) come from. The first is determined by the district and depends on available funding; the second is determined by examining the evaluations that have been done by the National Science Foundation and districts themselves on workshop programs that have been offered over several decades. The shape of the curves for the latter two variables was determined using these evaluations and through discussion with practitioners. The field

has developed a good approximation of these values. Interestingly, the content of the workshop does not matter. It is assumed that the designers incorporate the most effective content. Since the value for baseline effectiveness can be changed, one can easily determine the impact of greater or lesser content effectiveness on the workshop outcomes. While the functions for effectiveness and resistance I have used are estimates, an experienced planning team can produce a good estimate of the general shape of each function.

3. As a final step, I incorporated a linked stock/flow system that tracked program costs using stipends and fixed costs as detailed by the district.

Running the simulation only takes seconds, yet the insights can be profound. What's more you can then test those insights by changing the behavior of those variables (ie your assumptions), and the relative scale of those variables, and then watch what happens. Thus insights build on insights.

This SDSC yielded several important and non-intuitive results. The most important was that these workshops alone cannot adequately deal with the problem of building the necessary capacity in the teacher workforce. Even after 10 years of providing three-week workshops (a typical program), only slightly more than 50% of the teachers use reform-teaching methods – and this number includes teachers already in the system who were capable of doing so before they enrolled in the workshops and new teachers who entered the system with that ability. The results clearly showed that the workshops do not produce a critical mass of teachers with the desired capabilities. Even if ten years was long enough, this time frame is not acceptable for implementing so necessary a change in a district. That's a critical evaluative judgment, but even more important judgments are possible.

For instance, a somewhat unexpected result of this analysis is that although the number of teachers willing to enroll is significantly less for a long workshop than for the shorter workshops, the longer-length workshops result in the largest number of trained teachers over the ten-year period. This result highlights the strength of computer modeling; it allows us to examine the impact of input variables that produce a nonlinear response in outcome variables. In this case, the shape curve that tracks the relationship between the length of workshop and participant resistance to enrolling is different from that between the length of the workshop and its effectiveness in changing behavior. Thus, the percentage of teachers who change teaching behavior after the longer workshop more than offsets the loss of participants. Finally, the longer workshop is also the most cost-effective per changed teacher. This is a valuable insight that could change the overall evaluative conclusions about this program.

The model can be generalized since, as discussed previously, the results are independent of workshop content. The values of the quantitative variables are available (ie, number of teachers in the system, distribution by length of service,

teacher leaving rate, funding available for workshops) and the values for the qualitative variables (ie, workshop effectiveness, relationship of workshop length to teacher resistance and workshop effectiveness) can be reasonably well estimated from a district's prior experience.

This SDCS can also be easily extended in several important ways. One such extension is provided by the flow labeled *reform decay* that originates from the *Reform Teaching* stock. This flow is meant to illustrate that without continuing support during the school year some of the teachers will give up on reform teaching. Following up on this, we could extend the model to include the impact of resources provide and student achievement on this decay rate and, ultimately, on the accumulation of *Reform Teaching*. We could also examine the impact of rationing workshop participation depending upon the teacher's average time of service in the system, addressing the question "Should a system concentrate on those who will remain in the system longest?" Again it is unlikely that this would be picked up an evaluation of the program without using system dynamics modeling. Nor would it be possible to explore deeply the impact of that evaluative judgment.

The insights gained by analyzing this model of a program of professional development workshops illustrate the power of SDCS in understanding the relationship between a program's structure and its effectiveness. For instance, without the SDCS an evaluation of the program might support the decision to run short workshops because it increased the number of participants. However with the SDCS analysis this decision might be evaluated much less favorably because in the long term it reduced the impact of the change process.

Even as a first step, building an understanding that elements of a system interact through feedback loops, that causes often are not linked directly in time with effects, and that delays can cause unanticipated behavior are extremely valuable for both program designers and evaluators. The ability to capture nonlinear dynamics, identify unanticipated consequence of system behavior, and rapidly test policy options and resource allocations can make SDCS an important tool for many evaluation studies. Most SDCS studies to date have been in examining the function of business and other organizations. I believe that this method has great utility for evaluating social programs and that its increased use will benefit these programs.

References

- Forrester, J W. 1961. *Urban Dynamics*. Cambridge MA: Productivity Press.
- Richardson, G P. 1999. *Feedback Thought in Social Science and Systems Theory*. Waltham MA: Pegasus Press.
- Sterman, J D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston MA: Irwin/McGraw-Hill.
- Sterman, J D. 1991. A Skeptic's Guide to Computer Models. In, Barney, G O, et al (eds), *Managing a Nation: The Microcomputer Software Catalog*. Boulder CO: Westview Press. 209–229.

Other Resources

- The System Dynamics Society, <http://www.systemdynamics.org/>. This site will provide basic information and references to studies on the use of SD to evaluate many different types of social programs.
- The MIT System Dynamics Group, sysdyn.clexchange.org/sd-group. Dr Jay Forrester, MIT, pioneered the field of SD and he and the SD group continue to make many valuable contributions to its development and use. Of particular value at the site is the Road Maps series, a self-study guide to learning system dynamics.
- The Creative Learning Exchange (CLE) website, www.clexchange.org. While primarily organized to promote the use of SD in K-12 education, CLE is a very useful source of SD training materials and SD models.

USING SYSTEMS CONCEPTS IN EVALUATION

A Cybernetic Evaluation of Organizational Information Systems

Dale Fitch, Ph.D.

Although this chapter is ostensibly about evaluating management information systems in not-for-profit agencies, the potential application of Beer's Viable Systems Model in evaluation stretches way beyond this. Beer was interested in what helps and hinders effective communication within and between levels of an organization and its environment. And which evaluation isn't interested in that? Indeed, evaluation itself can be viewed as a form of management information system. Dale's chapter also usefully provides a heuristic of evaluation questions that can be asked at each level of the cybernetic analysis. This creates a valuable means of translating an unfamiliar model into the kinds of questions that will be very familiar to evaluators.

Introduction

Management information systems have quickly proliferated all aspects of organizational life. With the sums of monies invested in these systems and the amount of time and energy they require to maintain and operate, they naturally avail themselves to being evaluated. I will outline how cybernetic management principals, such as Stafford Beer's (1985) Viable System Model (VSM), can be used in evaluating an agency's systems of information with the aim of improving design.

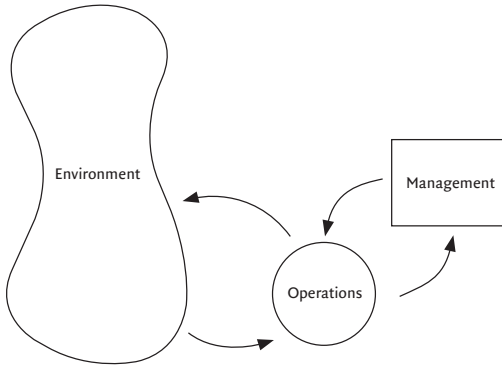
Specifically, I examine how an agency's system of information supports the adaptive connectivity of the organization (its ability to process information in its efforts to maintain existence). I have organized this chapter around the Viable System Model as proposed by Beer and discussed by others (Espejo & Gill 1997; Hilder 1995; Leonard 1999; Leonard 2000). I begin by identifying the System-in-Focus and proceed to cybernetic processes including viability, homeostasis, requisite variety, variety attenuation, variety amplification, entropy, negentropy, and invariance. The evaluation question is thus: *is the organization's system(s) of information adequate to manage the complexity, or variety, of the situations it faces in its day-to-day operations?*

The Viable System Model

Beer applied cybernetic principles to organizational functioning due to his appreciation of the complexity of forces involved in the organizational enterprise involving people, technology, information, etc. He wanted a framework that would allow a manager to manage this complexity for organizational survival. Therefore, Beer took the physiology of the human body's nervous system and applied it to the structure of an organization if it wanted to be a viable system, hence his viable system model.

Figure 1 provides the basis for the VSM. The Operations circle and Management box together would represent the body as they negotiate the Environment. Operations would be analogous to our limbs and their accompanying sense organs connected to Management as the central nervous system. As Operations encounters events in the Environment this information is transmitted to Management which provides the signal to control the actions of Operations interacting with the Environment.

Figure 1: A Basic Viable Model System



For the purposes of this chapter, I will begin modeling what this encounter would look like in an agency that runs a homeless shelter. Figure 1 is modified in Figure 2 to capture what these encounters and signals may entail.

Figure 2: An Applied Viable System Model

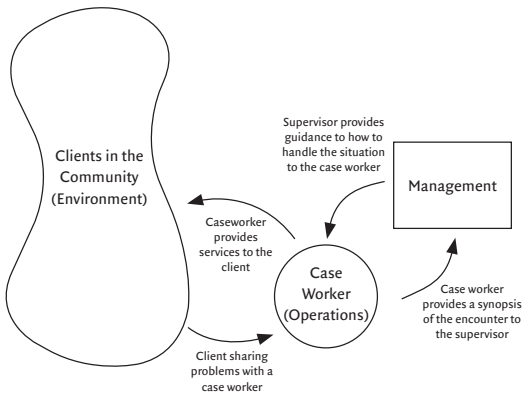


Figure 3: Viable System Model for the Shelter

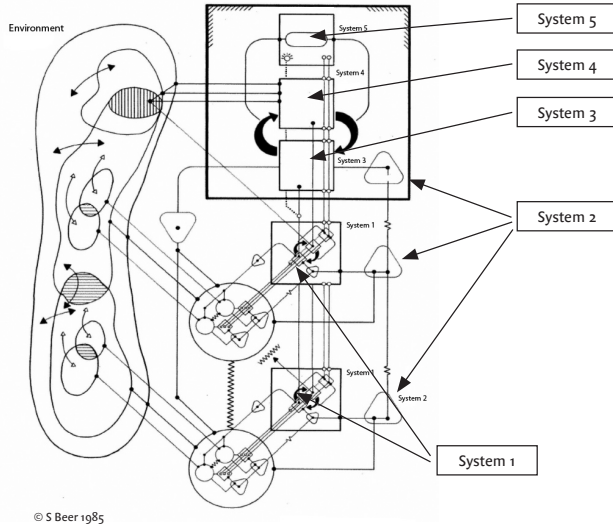


Figure 3 represents what the entire agency with all of its programs might look like.

As such, the System-in-Focus (Systems 1–5)¹, represents the entire Shelter with all of its programs.

In very general terms System 1 is *what* the organization does; System 2 is what glues the System 1's together (eg co-ordination). System 3 is *how* it does it what it does (ie management) and System 4 is *why* it does it (ie strategy). System 5 is the formal link with the environment (ie ownership) and together with System 4 ensures that the organization continues to do what it is supposed to do, or needs to do.

The two System 1s may be viewed as two programs within the shelter. One System 1 may be a job training program and the other System 1 a mental health counseling program. The smaller circles within the larger circle represent all the program staff. System 2 will be discussed later. System 3 would be conceived the first level of upper management and is typically occupied by a Program Director position. This system also performs quality assurance, human resources, and other functions. System 4 would be occupied by the Executive Director of the organization and System 5 would be the Board of Directors. Designating the Environment as everything outside the front door of the agency is important because it represents a boundary between the organization and the rest of the world through which

1 Following Beer's example, the following notation will be used throughout this discussion. A reference to System 1–5 should be read, "System 1, System 2, ... , and System 5." Likewise, a reference to System 3–5 should be read, "System 3, System 4, and System 5."

information must travel. The smaller ovals within the larger oval will be explained later. Most importantly, all lines connecting these systems and the organization (System 1–5) with the Environment represent lines of communication. And while drawn as single lines, they actually represent the two-way communication processes, feedforward and feedback, as illustrated in Figure 2.

Methodology

While I will use a hypothetical homeless shelter throughout this chapter, my discussion is based upon preliminary findings from my ongoing research project, System Design Modeling the Information Requirements of the Human Service Organization. This case study involves human services agencies in a mid-Western state of the United States and comprises a residential treatment center, an advocacy agency for students with school problems, a relief organization, a homeless shelter and a housing agency for senior citizens. Staff sizes range from six to over 100. Data sources include interviews with agency staff (individual and focus group), review of agency forms and paperwork including case record reviews, and observation of staff in their work environment including intake assessments, informal conversations with clients, and staff meetings. Interviews were only semi-structured depending upon the situation, but typically included the following questions:

1. What information do you need and use to perform your job?
2. What information do you use when making decisions?
3. Where does this information come from? / How do you get it?
4. Where is information about clients kept or stored?
5. Who else looks at this information in addition to you?
6. Who else needs this information?
7. Is the information you gather information you would routinely want to share within the agency?
8. How does this information sharing currently happen?
9. Where do you record information that helps guide your decision-making processes?

System One

The square in System 1 represents the supervision of what happens in the circle which is the operational function of the system (eg the management of staff dealing with employment issues with clients or the management of staff dealing with mental health issues with clients). These clients originate from their respective ovals in the environment. The area where the ovals overlap may be viewed as clients who have mental health issues that impacted their employability (or vice versa). The line between the ovals and their respective circles in the Environment represent the exchanges of information that occur when a client seeks services. Lines from the circles to their corresponding squares represent information communication from staff to their supervisor and is needed by

management to ensure that appropriate services (amount, type, staffing, etc.) are being provided to clients. In total, these lines represent the management of complexity or the *variety* the organization must be able to handle in order to insure that the issue of homelessness in the community is dealt with properly ².

The amount of complexity presented by the environment (the collective complexity in the lives of those who are homeless) is higher than the complexity of the staff in the operational circle which, in turn, is higher than management. In fact, some agencies refer to their Intakes as a “screening” process – literally, only information about the client that is pertinent to what the agency can address is recorded. In turn, staff provide a synopsis of the clients with whom they have completed intakes with their supervisor for staffing purposes. This process of paring down information is cybernetically known as *attenuation*. In response, and in order to maintain homeostatic balance (ie maintain internal stability), the organization amplifies its variety (ie *variety amplification*), by hiring qualified staff and offering programs based upon research in the given area. Balancing the amount of variety that is attenuated and amplified, Beer asserts, is what management is all about and it is done to insure organizational viability.

Beyond the programming and staffing are lines of communication that make the whole process work. Communicating this information occurs verbally, via the agency’s paperwork procedures and/or a MIS. From the outset it must be stated that in no way will a MIS ever be able to serve as the only means to communicate information in an organization; life simply has too much complexity. At this point, the evaluation questions will become more specific and may encompass:

- to what degree can the MIS serve as a tool for communicating information and, more importantly,
- in what ways might it be more efficient than voice or paper (analog processes), as a means to communicate this information for a particular agency?

Cybernetically, this would be the degree to which the agency’s systems of information are able to maintain homeostatic balance both internally and externally between the organization and the environment.

For example, a 20-page intake form would undoubtedly have more information about a client than a 2-page intake form. While a longer form will produce more information it will also take longer to complete and may not result in any different services being provided. On the other hand, agencies with relatively shorter forms are almost always augmented with personal notes written in the margins and elsewhere in the client’s chart. This latter reflects variety amplification, ie the existing form did not provide enough places to record notes so staff added information to facilitate

2 This last statement could conceivably raise a whole host of evaluation questions about whether that is happening or not and the degree to which the services are effective, etc. Those types of evaluation questions are addressed by other chapters in this book. A given in this situation is that the agency has the proper programs and is doing the best it can with what it has.

their decision making. The problem with the method employed, however, was that it was idiosyncratic among staff with no two doing it the same way and with some of the information being 'lost in the chart' resulting in a net loss of information.

An additional critical factor when looking at the process is the amount of time involved. Time on task is an overlooked factor in some evaluative methodologies, but it plays a key role in understanding the cybernetic functioning of an agency. A 2-page form is shorter and takes less time to fill out.

- But do staff spend the extra time appending the form?
- If so, how much time is spent trying to locate information in a chart that may be found in any number of places?

A 40-page form will take longer to fill out.

- But is it the right amount of information or is too much information gathered that nobody ends up using?
- Are there a lot of blank areas in the form meaning that no one is obtaining this information in the first place?
- And is the information used? In the diagram, this might mean lines from system 1 that reach a dead end in System 3.

In observation, case study agencies displayed several different behaviors in compensating for the shortcomings in their information systems when the demand for services was high. Beer would predict such behavior due to his assertion that organizations instinctively attempt to manage the complexity for which they are responsible. For example, it is a busy Monday morning with several intakes that need to be completed first thing. Forms are filled out, but only with the basic information. When it is time to determine eligibility for services, the written information is augmented by verbal information. On an occasional basis this type of behavior is not necessarily problematic. However, if this occurs frequently then it would seem that the agency's paper system of information has insufficient *channel capacity* (ie the ability to transmit needed information in a given amount of time), resulting in the use of verbal communication. Again, this is not necessarily problematic except that verbal communication has to be synchronous. Furthermore, it was common to find this verbal communication augmented by a note, typically of the yellow "sticky note" variety, attached to people's chairs or computer monitors. The net result is that client information is now located in the chart, transmitted verbally, and captured in an unofficial note.

This redundancy (ie excess effort) brings us to the phenomenon known as *entropy*. Entropy can be understood in terms of thermodynamics, statistical theory or information theory (de Rosnay 1998). Regarding the latter, Ashby (1963) used Wiener's (1948) arguments to demonstrate cybernetically that information is *negative entropy* in that according to communication theory it produces results. Stated otherwise, data that provides information is negentropic; data that does not provide information is entropic. Granted, there is a debate about Wiener's conceptualization of entropy (Corning 2001); regardless, staff who spend their

time documenting in a client's chart, leaving a phone message, and writing a sticky note are engaging in redundant behavior that cannot be considered cost effective, unless this is the only way to let the recipient know the urgency of the message. As such, an evaluation question would be:

- *Are staff engaged in multiple methods of communicating information due to inadequate formal channel capacity? If so, what can be done to reduce the need for this redundancy?*

While all of the above processes (variety attenuation, variety amplification, channel capacity and entropy) have been discussed within the context of System 1, they also occur within all the other systems to be discussed.

System Two

Unlike the other systems with which people are typically associated, System 2 captures the policies and procedures that keep all the System 1s operating smoothly. For example, there is most likely a procedure in place to work with a homeless person who has both employability and mental health issues. The relationship between the two programs is noted by the serrated lines connecting the System 1 operational circles. Keeping these relationships regulated is the function of System 2 and is typically manifested by scheduling or calendar functions, documentation procedures, program guidelines, etc.

In my research I have chosen to focus on the information management capabilities that should fall within the System 2 domain. For example, if in the job skills System 1 program a staff person is working with Client A, then System 2 would also let them know that a staff person in the mental health System 1 program worked with this client last year. The knowledge of this prior contact would be discovered in a chart review so a typical System 2 policy and procedure would look something like, "Please review the client's chart on all new intakes to determine if the client has had prior contact with the agency." This procedure works fine if at least three conditions are satisfied: 1) chart notes are promptly written, 2) chart notes are filed in an orderly fashion, and 3) the client's chart is readily available. Ideally, if the clients were in a MIS, then the staff would know at a glance if a client had been served previously by the agency because access would not be site specific and the information is always in the same place in the database. To evaluate the cybernetic qualities of this aspect of System 2 for an agency, the question might be formulated as:

- *Does the agency's system of information provide or deliver information about prior client contact to staff or is it expected of a staff person to track down a physical chart?*

Then there is the situation of the client who is referred between programs with appropriate releases of information having been signed. If the documentation on client progress is done in the same chart and if the analog chart is readily available, then it is possible for the program that did the referring to know if they had made

an appropriate referral to the other program. However, if the referring program receives no feedback, then they are essentially making blind decisions with unknown effects upon the homeostatic balance of the organization. The evaluation question for this type of phenomenon would be captured by:

- *Does the agency's system of information provide feedback on the appropriateness, or even outcome, on all referrals made?*

One final example of cybernetic control exhibited by System 2 is the ability of the agency's information system to help staff manage client caseloads. As previously stated, none of the agencies researched have a MIS that manages client data for System 1 purposes. As such, most of the staff work with paper charts, file cabinets, etc. When asked how they manage what they need to do with their clients they invariably open their personal notebooks and produce a "Things To Do" list. This situation is ideally suited to be addressed by a MIS solution and will be discussed under Implications. For evaluation purposes the question would be:

- *Does the agency's system of information facilitate case management activities?*

System Three

Beer sees System 3-5 in the job of managing the complexity generated by all the System 1s engaging with System 2 to keep all this System 1 generated complexity – amplified and attenuated – in some type of balance with the environment, ie the community. In order to keep balance, upper management must know what this complexity entails.

We have already discussed that System 1 largely has to work with analog information systems (paper and pen, voice) to record and transmit the vast amount of complexity they encounter. To gain a sense of this complexity, Beer asserts that System 3 implements "audit" functions to make sure all the System 1 operations are being conducted within the resource constraints of the organization. This function is typically carried out by requesting either weekly or monthly statistical reports from the System 1 staff consisting of "total number of clients served," "average number of clients per staff/program," etc. This is where things get complicated (cybernetically speaking). To produce these statistics, System 1 staff fill out tally sheets derived from reviewing their case files, phone logs and other information. As such, all the work with a client may be boiled down to a single number, a case, and maybe what services were provided to the client. That's all! We now understand that in gathering this type of statistic a tremendous amount of complexity has been summarized. Granted, System 3 does not need to know every little detail concerning a client, only those things they believe will assist in maintaining balance for the organization. But I have never encountered a situation in which System 1 staff, and System 3 staff included, felt that these monthly statistics capture the work of the organization.

This process provides the basis for the following evaluative questions:

- *Does the agency's system of information facilitate monthly statistical reporting?*

- *How much complexity is lost by the current system of information employed?*
- *Is the monthly statistical reporting method employed a net entropic process?*

A feature common throughout the agencies under study was that, unlike the System 1s, the digital information systems available to System 3 were ubiquitous throughout all the agencies regardless of size. The lowest common denominators were spreadsheet applications used by System 3 administrators to tally the requested numbers. While not necessarily a purview of cybernetic processes, this observation is not lost on the System 1 staff. Whether true or not, a matter of value or worth is sometimes associated with being able to use computers or information technology in the service of one's job. Whether it be statistics, budgeting, fundraising, etc, a whole host of digital tools have been developed for System 3, and there is absolutely nothing wrong with that. However, if the purpose of what an organization does is carried out by System 1, then shouldn't System 1 have access to the same information system tools afforded the other Systems in the organization?

Secondly, System 3 is usually charged with writing the policies and procedures related to information system security carried out through System 2. Typically line staff, primarily not professionally educated, are not given access to an agency's MIS due to concerns about maintaining the confidentiality of client information. However, as was discussed with System 1, these are the very staff who have the most contact with clients in many agencies. Having to deal with this constraint imposed by System 2 is not uncommon in organizations. Identifying these system barriers almost always leads to discovering organizational information structures that impede optimal cybernetic functioning. As such, the evaluation question would be:

- *What workarounds do the organization employ in being able to capture or transmit important information relative to client care?*

Another phenomenon was encountered in those agencies that had MISs that System 1 staff could use. While these systems were designed to be capable of recording client information, the reporting features available in the database were primarily designed to meet the information needs for System 3-5. Typically, these needs are defined by the Environment at this level (see lines connecting System 3-5 with the Environment in Figure 3) and are reflected in reports to funders, either governmental or foundations, and have to be produced on a quarterly to annual basis. As such, the MIS is viewed by System 1 staff as only having a "reporting out" function which means data entry is almost never done in real time, sometimes over two months after a client has received services. While issues of data quality arise in these situations, the most important concern is what data are staff using in making decisions in the meantime? These issues lead to the following evaluative questions:

- *If the agency has a MIS, do operational staff enter data on a real time basis?*
- *If not, what is (are) the source(s) of the information used by staff for decisions that need to be made in System 1?*

- If not, how does the time lag affect decision making at System 3–5.
- On balance, are agency reports primarily generated for entities outside the agency or for internal staff?
- Do Systems 1 and 3 answer the following questions differently: What information do we gather? What information do we really need? If so, in what ways?
- Do the data elements in the MIS allow the agency to determine whether or not it will be able to achieve its program goals and client outcomes?

Finally, possibly the cybernetically worst use of digital data involved those situations in which reports produced by a MIS were via a static document format (eg PDF). This situation is not necessarily problematic as long as the reports produced answer every question a person might have about the operations of the organization. Unfortunately, most often it was a case of someone asking to look at the frequencies or averages in a slightly different manner necessitating the running of another report, configuring a new ad hoc report, or, worse yet, taking the numbers from the static document and entering them into a spreadsheet. These three scenarios all require additional time which slows the feedback loop of the decision making time cycle. This type of organizational behavior can be assessed by inquiring:

- If additional reports are required, is it truly for an ad hoc situation or has some organizational function been systematically overlooked when assessing reporting requirements?
- In what types of formats are reports available?
- If additional data manipulation is required, does it result in a time lag for decision making and, if so, for how long?

System Four

While previously alluded to in the discussion up till this point, Beer posits that if the focus on System 3 is on the *inside-and-now*, then System 4 has a focus on the *outside-and-then* as its job in maintaining organizational viability. System 4 is the face of the organization to the larger community (Environment). Its job is not only to provide overall leadership and management of the organization, but also to link the organization with the environment in terms of securing resources, specifically funding, to maintain year-to-year operations. The ovals in the upper half of the Environment in Figure 3 may then represent state or federal government entities, private interests, etc. System 4 is expected to establish relationships with these constituencies much the same way that System 1 establishes relationships with clients. In context, System 1 is what the organization does; System 3 is how it does it; and System 4 is why it does what it does. Together with System 5, System 4 incorporates feedback from the Environment in developing the organization's Mission Statement and sets the course for achieving this Mission in strategic planning documents.

All of the cybernetic issues discussed thus far (variety attenuation and amplification, channel capacity, entropy, etc) culminate in the linkage between

System 3 and System 4. In many ways we are asking that this linkage convey all the complexity that is necessary to balance what is happening in System 1, via System 2–3, with needs in the Environment and as interpreted as to what can be done by System 4. The primary information systems to maintain this linkage and convey variety are the reports generated by System 3, of which their limitations we are now familiar, and verbal communication that occurs in agency committee meetings. With this mixture of digital and primarily analog information systems, the question is:

- *How much does our reliance on analog systems of information, either paper or person, more so than digital systems, eg MISs, affect or alter how well System 4 is able to perform.*

This question is best answered by reverse engineering reports given to System 4 to find out from where they originated then determining at which points the information existed in either digital or analog form or both. Where they existed as both, then the organization knows that wasteful processes are in place.

In an age when we are supposedly operating in information overload or “garbage in – garbage out,” I have yet to meet a System 4 human services executive who says that she or he has too much information about the agency they are expected to run. Granted they can be inundated with data, but not information gleaned from such data. Such is not the case in the corporate sector where Executive Information Systems (EIS) have been around since the early 1980s (Kaniclides & Kimble, 1994) to help chief executives manage internal and external information necessary for strategic planning (Executive Information Systems, 2005). Unfortunately, since none of the agencies in the current study have anything that even resembles an EIS, the discussion of the information processes at this level largely echo what is occurring at System 3 except System 4 has more direct benefit from System 5.

System Five

This system represents the function of the board of directors for an organization. Comprised of members from the community (ie the Environment), their role is to increase organizational intelligence and to provide leadership and serve as a resource to System 4 to make sure the organization will have some type of programmatic and economic viability five to ten years in the future (if possible).

Together with System 4, System 5 is responsible for determining whether or not the agency is doing what it is supposed to be doing *a la* its Mission Statement. The answer to this question is typically found in an agency’s Annual Report. Ideally the information contained in the report is derived from digital information sources that have been refined to provide a coherent picture of organizational functioning. This functioning builds on previous questions to lead us to ask:

- *Do the data elements in the organization’s systems of information allow the agency to determine whether or not it achieved its Mission?*

- Does the Board of Directors make use of aggregated client data in their discussions with System 4?

There is one final evaluative question that can and should be asked at every level of recursion in the organization, Systems 1 through 5. Beer maintains throughout his discussion of the VSM that, ‘the purpose of a system is what it does’ (p. 128). For our purposes, the converse could be very informative:

- What can we deduce to be the purpose of this agency by analyzing its system(s) of information?

Again, an organization’s systems of information will not convey every nuance of complexity necessary to carry out its function, but an analysis of its information systems should give you some semblance as to what it is about.

In conclusion, I want to return to Figure 3 and to what appears to be a version of System 1–5 residing in System 1. Indeed, that is exactly what it is. If we were to look down one level of recursion from the mental health program in the shelter, and if this program were rather large with several professional staff, technicians, office assistants, etc, then it would have the same operational issues concerning information flow that affect the whole organization. This is a manifestation of invariance endemic to systemic behavior. Now, envision that the whole System 1–5 we have been discussing is actually residing in its own System 1. Such would be the case if the shelter, or the System-in-Focus, were one of several shelters in a community existing in a network to meet the needs of the homeless population in a large metropolitan area. Again, the same informational flows would be at work and the evaluative questions would be no different.

Implications for MIS Design

In many ways this cybernetic evaluation can perform part of the requirement analysis process in designing a MIS for an agency. As an iterative process, this evaluation can be done prior to designing and installing a system or, more likely, after a MIS is already in place to determine what specific aspects of its design need to be modified for cybernetic efficiency. As discussed here, modifications may not need to be made with the MIS itself, but rather with the policies and procedures governing its use or even possibly a hybrid digital/analog solution. For example, regarding the agency that prevented line staff from accessing the MIS, all staff were given access to the MIS but only to a portion of the client record. For other information they needed to know, a query/report was created that pulled information from the database and it was printed out for particular staff meetings, a variety amplification process. So, even though an analog process was employed, since it was automated via the database, having that information was felt to be more important than not having it so the net effects were negentropic.

While the reporting capabilities of a MIS for System 3–5 will most likely continue to take precedence, its digital capabilities can also have added benefit

for System 2 purposes. As mentioned earlier regarding intake forms with blank responses, almost all of the agencies have a quality assurance process in which they do nothing more than review charts to make sure all the blanks are filled in. With database technologies, the same task can be handled in milliseconds by doing queries on fields with null values. More importantly, System 3 can review these null value results and inquire with System 1 why certain fields are always blank. System 3 may learn from System 1 that the population has changed and that it may be preferable to redefine that particular data field to better capture what is going on with clients, ie enhancing the intake form's ability to capture the complexity originating from the Environment.

Additional Subjects for Cybernetic Analysis

While the scope of this analysis was restricted to the organization's information system, it is important to remember that the purpose of this information system, specifically the information in the system, is in the performance of the work of the organization. This information thus allows the organization to perform its work and the degree to which it does so should then be fed back into the (information) system so as to complete the circuit or close the cybernetic loop. The degree to which it does so would, in itself, be a good subject for further cybernetic analysis.

In addition, one could examine an organization's entire systems of communication. These systems would include meetings, memos, emails, online collaboration tools, water cooler or coffee pot conversations, their World Wide Web presence, etc, to analyze how communication channels are utilized as a gauge of organizational system functioning.

Indeed, Beer asserted that managerial cybernetics could be applied to a whole range of subjects and a review article by Vancouver (1996) listed the following studies: adaptation to work transitions (Ashford & Taylor 1990); decision making in humans and organizations (Beach 1990); stress in organizations (Edwards 1992); emerging organizations (Katz & Gartner 1988); and leadership and information processing (Lord & Maher 1993), to name just a few.

Final Thoughts

The decision to install or customize an existing MIS is not cost inconsequential either at time of purchase or in ongoing maintenance. Indeed, a possible outcome of conducting a cybernetic evaluation of an organization's systems of information may be the decision not to purchase a MIS. However, rarely do our organizations evaluate their analog systems of information in terms of the same costs with staff time and information lost. When the wasteful processes associated with managing analog information are added on, then the benefits of a MIS might start to make sense.

References

- Ashby, W R. 1963. *An introduction to cybernetics* (Science Editions ed). New York: John Wiley.
Available for download from the *Principia Cybernetica Electronic Library*,
<http://pespmc1.vub.ac.be/LIBRARY.html>
- Beer, S. 1985. *Diagnosing the system for organizations*. New York: John Wiley.
- Corning, P A. 2001. "Control information": The missing element in Norbert Wiener's cybernetic paradigm? *Kybernetes*, 30(9/10): 1272–1288. Also retrieved August 3, 2005, from http://www.complexsystems.org/publications/pdf/control_info.pdf
- de Rosnay, J. (1998). Entropy and the Laws of Thermodynamics. In, *Principia Cybernetica Web* (Principia Cybernetica, Brussels), Heylighen, Joslyn F C, and Turchin V (eds). Retrieved July 30, 2005 from <http://pespmc1.vub.ac.be/ENTRATHER.html>
- Espejo, R and Gill, A. 1997. *The viable system model as a framework to understand organizations*. Retrieved June 20, 2005 from http://www.phrontis.com/acatalog/Free_Papers.html
- Executive information systems (2005, March 22). *Wikipedia: The Free Encyclopedia*. Retrieved July 10, 2005, from http://en.wikipedia.org/wiki/Executive_information_systems.
- Hilder, T. 1995. *The viable system model*. Retrieved June 28, 2005 from <http://www.users.globalnet.co.uk/ffirxv/orgmgt/vsm.pdf>
- Kaniclides, T and Kimble, C. 1994. *Executive information systems: A framework for their development and use*. Retrieved July 9, 2005 from <http://www.cs.york.ac.uk/ftpdireports/YCS-94-247.pdf>
- Leonard, A. 2000. The viable system model and knowledge management. *Kybernetes*, 29(5/6): 710.
- Leonard, A. 1999. A viable system model: Consideration of knowledge management. *Journal of Knowledge Management Practice*, 1(1998–1999), July 1, 2005. Retrieved July 1, 2005, from <http://www.tlinc.com/articl12.htm>
- Tellis, W. 1997. Introduction to case study [68 paragraphs]. *The Qualitative Report*, 3(2), July 1, 2005. Retrieved July 1, 2005, from <http://www.nova.edu/ssss/QR/QR3-2/tellis1.html>
- Vancouver, J B. 1996. Living systems theory as a paradigm for organizational behavior: Understanding humans, organizations, and social processes. *Behavioral Science*, 41: 165–204.
- Wiener, N. 1948. *Cybernetics*. New York: John Wiley.

Soft Systems in a Hardening World: Evaluating Urban Regeneration

Kate Attenborough

Today, multiple perspectives are an established norm in many evaluations and systems-based initiatives. So it is hard to grasp how revolutionary and innovative soft systems ideas were when introduced in the early 1970s. Yet they marked the departure from systems being seen as essentially descriptions of the real world, to being seen as conceptual ideas that help us learn about the real world. Soft systems has been hugely influential and the basis for many thousands of interventions over the past forty years. It was also a vital stepping stone to the next “wave” of systems concepts – critical systems thinking. It also cemented – or made explicit – the close relationship between systems concepts and action research concepts.

Kate’s case study is a classic example of when soft systems methodology (SSM) has been used. These are complex, messy, community or institutionally based situations capable of multiple points of view. Clarity, and learning arises by untangling these perspectives, considering them one at a time and comparing each with the real life situation.

As you read this case bear in mind the following. Multiple perspectives are quite different from multiple stakeholder views (different stakeholders can share the same perspective, single stakeholders can have several perspectives). Soft systems methodology is an exercise in both logic and inspiration – it is far more rigorous than it appears to be. Although Kate’s example covers an entire large scale intervention, SSM can be used small too – perhaps as a desk study at the design stage of an evaluation to work out appropriate questions or where to place the boundary of the inquiry.

1. Introduction and background

This chapter describes how I have used soft systems methodology (SSM) in evaluation in the community sector. It gives an overview of the methodology, illustrating this with one particular application of the methodology and concludes by summarising the benefits of the approach, in particular for the community and voluntary sector.

I use an informal style for three reasons: to acknowledge that subjectivity is present in all we do; to reflect the need to encourage participation and knowledge sharing through employing jargon-free language; and it is consistent with Eden et al (2005) who suggest that words used by problems owners should be kept unchanged because this represents their own perceptions and creates senses of ownership and accountability.

2. Overview of Soft Systems Methodology (SSM)

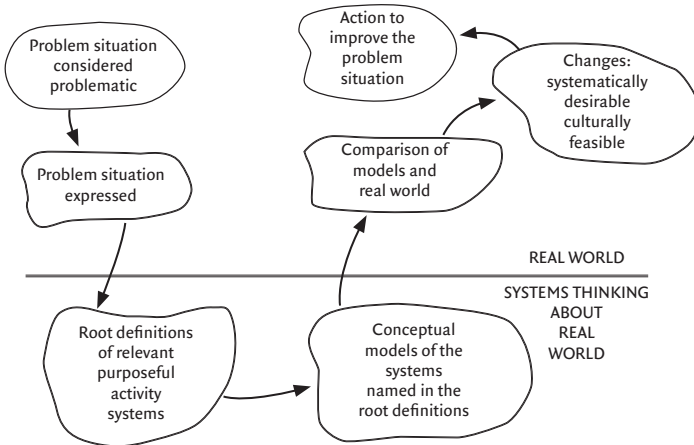
Peter Checkland developed soft systems methodology in the late 1960s at the University of Lancaster in the UK as a way of acknowledging that systems don’t

define themselves, people define systems. And since people share different views of a situation many systems can be defined within that single situation. So systems approaches need to be able to model purposeful activity while encompassing multiple viewpoints. This marked a departure from purely describing systems as they are in the real world either by constructing models (eg system dynamics) or by biological analogy (eg cybernetics). To account for multiple perspectives Checkland added a stage to a systems inquiry that helped people develop ideas of how systems might be conceived in a particular situation and how to use the insights gained to make improvement to the real world.

In other words he separated out the “real world” from “systems views of the real world” (see Figure 1 below). Learning occurred by comparing the similarities and differences between these two.

There are many variations of SSM, but the classic version comprises seven stages (Checkland & Scholes 1990). Five relate to the real world and two comprise systems thinking about the real world.

Figure 1: The classic soft systems methodology



In my own words and adapted for ease of explanation to colleagues, the seven stages are (working anti-clockwise from top left):

1. The situation “as is” – the “mess”, the problem situation defined
2. The situation expressed (eg “rich picture”, list of issues, tasks and components in the situation)
3. Root definitions of relevant systems (the essential components of the conceptual models)
4. Conceptual models (ie the activities that are logically necessary to achieve the particular purpose of the system defined in the root definitions)

5. Comparison of models with reality (stages 4 and 2)
6. Options for change – that must be both desirable and feasible
7. Agenda for change.

Expressed this way the process appears very sequential and linear. In practice, the process is iterative (as a whole and in part) and I often revisit many stages many times. Indeed Checkland strongly recommends that you do so. Whilst the separation between the “real world” and “systems world” is critical, the distinction between the stages is not precise. As you will see below, I use a slightly different approach to the first two stages.¹

3. Using SSM in regeneration: an example of formative evaluation

My first job in the community sector was in Darnall, a neighbourhood in the city of Sheffield, England, where a major urban regeneration scheme was nearing completion. As a manager with an area-wide remit for European Community funding which was to follow, I had to quickly take stock of what had been achieved already, as the new programme would start before formal external evaluation of the previous one started. I had used SSM many times and thought it could be helpful here.

STAGE 1: THE SITUATION “AS WAS”

The first step is to explore and define the “problem situation”. Checkland developed the methodology to solve problems, hence the term, but as Bob Williams observes it could be equally “program”, “issues” or the kinds of words we use in evaluations, (Williams 2005)². Checkland often uses the term “mess” to describe the range of contradictory views and experiences when seeking to define the problem situation. Within my first week, I found satisfied funders and developers, a community forum (one of two created by the authorities because the first one did not work as intended), committed and vocal community activists, disenchanted residents who had not benefited from the developments, and my employing organization perceived as a threat by two local community forums.

Checkland suggests collecting as much data as possible, using whatever methods and sources are available: “survey, observation, measurement” (Williams 2005). In practice, this gives me the best possible chance of deriving a sound analysis, and realistic options for change, or in this case for the next regeneration programme and its evaluation. I use business plans, performance information, gossip, memos, minutes, reports, the internet (for comparative data), statistics, survey findings, anything relevant that I can lay my hands on.

STAGE 2: PROBLEM SITUATION EXPRESSED

Checkland suggests using a picture to express the situation as fully as possible – the rich picture. Indeed, he comments that during the 30 years he has been

1 Boon Hou’s chapter is based on a radical modification of the SSM process.

2 Ken Meter makes a similar point in his chapter.

refining SSM, he has tried all kinds of ways to express the situation and finds that “picturing” seems to resonate best with stakeholders and participants.

I also find that using pictures and diagrams frees up my own thinking, releasing the constraints imposed by words, statistics and existing views and findings.

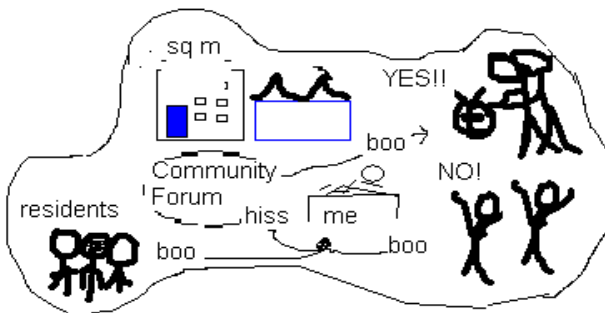
Checkland suggests considering the following aspects of the problem situation when drawing the rich picture :

- Structures
- Processes
- Climate
- People
- Issues expressed by people
- Conflicts

In Darnall, I spent a couple of days drawing bits and pieces of the situation “as was” – and drawing several different versions of the picture. There follows a severely edited version of my final rich picture, with only the barest essentials. I actively dislike drawing, yet find that the “rich picture” approach enables me to get to grips very quickly with everything that is important: structures, processes, stakeholders, views and opinions, issues and tensions, outputs, outcomes, and tasks.

This diagram (Figure 2) may suggest that a few quick sketches are enough. In fact my rich pictures are always based on all the information that I can pull together within 2–3 days.

Figure 2: The situation “as was” in Darnall, the area being regenerated



This stage of the process raised several important evaluation questions. In particular:

- Why were funders and authorities proclaiming the regeneration scheme an unqualified success when the people infrastructure was falling apart, and residents disaffected?*
- What were the tasks to be carried out and the issues to be addressed in any future programmes?*

In recent versions of the methodology Checkland recommends listing the key tasks to be achieved and key issues to be addressed. I find this useful and also then to share the lists with stakeholders so that I can check I am on the right lines. In Darnall, there were many key issues and several critical tasks and so I give only a short extract here. I have chosen the particular tasks and issues that were picked up in subsequent funding initiatives and evaluations. The word “task” does not mean that you are proposing “solutions” – this is a description of the key “tasks” observable from the situation. It is a step to help identify the key “systems” that you wish to model in the next stage of the methodology.

I identify the key tasks and issues in the system through looking at my rich pictures, thinking about them, adding to them and changing them from time to time over a few days, refining both stages. In Darnall, whilst the main “task” from the funders’ viewpoint was to physically regenerate the area, the idea had been presented to residents in terms of new houses and jobs – a second key task, and one that raised a key issue when the houses and jobs were not forthcoming.

Key Tasks

- Physically regenerate old steelworks area
- Develop community infrastructure/organizations
- Connect residents together
- Connect residents to employment opportunities

Key Issues

- People used as pawns
- “Cherry picking” of local representatives
- High cost regeneration “squeezed out” for lower cost regeneration.

STAGE 3: ROOT DEFINITIONS OF RELEVANT SYSTEMS

Next in the methodology comes one of the more difficult steps, for me anyway. This involves deciding a single perspective on which to base a systems view. Checkland recommends that you consider several perspectives to reflect the multiple perspectives present in any situation. For example a project may be seen from an “efficiency” perspective or an “effectiveness” perspective. SSM can help to clarify the implications of these different perspectives by identifying and developing systems insights and models from each in turn. Later, when dealing with conclusions and recommendations, an awareness of these different perspectives can help to ensure that future action is appropriate, or is able to accommodate contradictions between these perspectives.

This stage of the methodology involves developing “root definitions” of the different perspectives, so-called because “everything else grows” from (literally rooted in) this definition (Williams 2005).

A root definition is a short description of the system being modelled. Checkland uses the mnemonic CATWOE as a guide to help the construction of the

systems model (Checkland and Scholes, 1990).³

- Customers who are this system's beneficiaries
- Actors who transform these inputs to outputs
- Transformation from inputs into outputs
- *Weltanschauung* relevant viewpoints and assumptions
- Owner to whom the "system" is answerable and/or could cause it not to exist
- Environment that influences but does not control the system

In constructing CATWOE, I first list all the possible stakeholders to ensure that they are included as customers, actors or owners, and then ask myself the following questions:

- **C: Customers**
 - Who are the victims or beneficiaries? (eg clients, residents)
 - What are their experiences and views? Their needs and aspirations?
- **A: Actors**
 - Who will "do the doing", make things happen?
 - What are their experiences and views?
- **T: Transformation process**
 - What possible transformation processes are there (ie input to output) ?
 - What are all the steps in the process that transforms inputs into outputs?
 - What are the inputs, and where from?
 - What are the outputs, and what happens to them next?
- **W: World view**
 - Whose worldview are we talking about? Have I tried all possibilities?
 - What is my own worldview and what influence does it have here?
- **O: Owner**
 - Who has the power to stop the process or situation? Could this change?
 - Can the owner(s) help or hinder? (I try this with different possible owners)
- **E: Environment/constraints**
 - What are the constraints eg funding, legislation, time, power?

I find the only way with this stage is patience as I work my way through different possible definitions that reflect different perspectives, stakeholder viewpoints, key issues and key tasks. As recommended by Checkland I make a point of considering at least two root definitions in any piece of work.

Table 1 gives a feel for some of the perspectives present in the Darnall situation and lists components of six possible "systems" perspectives.

Checkland recommends that you construct these tables starting with the desired

3 It is worth noting here that in recent years, some associated with Critical Systems Thinking who use SSM have made two very significant changes to CATWOE (eg Martin Reynolds and Gerald Midgley in this volume). They have replaced C with two concepts; B for Beneficiaries, and V for Victims (BATWVOE) and B and V can include ideas as well as people. A nice example of a method developed during the "second wave" method being modified by "third wave" thinking.

transformation (T of CATWOE), and then working out what the appropriate customers, actors, owners, worldviews and environmental factors might be.

Table 1

	Customers	Actors	Transformation	Worldview	Owners	Environment/ Constraints
1	City politicians	City Council	Contaminated, derelict gateway to city and industrial land redeveloped Reduction in unemployment	Areas hit by steel industry decline need financial help Employment policy	Funders/ government	Available & potentially available funding European funding rules
2	Developers	Planners Construction companies	Contaminated, derelict gateway to city and industrial land redeveloped	Deliver the master plan, raise inward investment & help residents	Sheffield Development Corporation (SDC)	Lack of match funding from sources other than statutory agencies and government departments
3	National funders (and government)	Planners Developers Construction companies	Reduction in youth trouble	Pride in city's good reputation Local votes	City Council	
4	Residents (Darnall)	Statutory agencies Community Fora/groups	Unemployed steelworkers acquire new skills and jobs	Conflicting worldviews (planning v landlord v...)	Politicians	
5	Residents (Darnall)	Statutory agencies Community Fora/groups	Greater community control/transfer of some power to community	"Local jobs for local people" New Village	Community Forums	
6	Residents (S Yorkshire)	Statutory agencies Community Fora/groups	Reduction in unemployment across South Yorkshire	Strategic importance of zone to South Yorkshire	All Councils	

When using soft systems collaboratively with colleagues in community organisations (and previously in the public sector) I have kept to plain English: What systems are most relevant to the task and to us? Who does what to whom? What are they trying to achieve? How exactly are they trying to do it? Whose opinion wins the day? Who actually has the power and controls things? What are the barriers and limits that we have to work with?

Table 1 shows there can be many different relevant systems, depending on viewpoint and desired result. Two contrasting ones that illuminate the Darnall situation are:

A Redevelopment System (System #2 in Table 1)

(C= Developers, A= Planners, construction companies, T= derelict steelworks to development sites, W= investment/investment with benefits for residents, O= SDC, E= contamination, funding criteria/hectares improved/new jobs created)

Using these building blocks the “root definition” of this system, giving its purpose and key players, might be:

“A Sheffield Development Corporation-owned system which plans and monitors the physical regeneration of Darnall, using planners and construction companies to deliver specific projects according to externally-set regulations and in line with fixed financial limits, specific targets, and timescales. The system responds to community issues of employment and training, where these are consistent with funding and targets.”

A Community Control System (System #5 in Table 1)

(C= local residents, A= Statutory agencies and community organizations, T= increased community control, W= master plan and investment with benefits for residents, O= Community Forums, E= contaminated land, funding criteria/hectares improved/new jobs created)

So the root definition of this system might be:

“A system owned by Community Forums in Darnall which uses opportunities afforded by regeneration programmes to increase the degree of community control and transfer some power from the City Council to the community, through commissioning services from community organizations and statutory agencies working together or separately for the benefit of local residents.”

Some systems thinkers and practitioners emphasize the need for “consensus” definitions. This can tempt people to settle for compromise, and lead to a solution that displeases everyone equally all the time. I prefer to highlight choice and solutions to complexity rather than work towards a compromise that may resolve an issue but fail to achieve a critical task. Hence emerging options for change need to take into account movers and shakers who are passionate about what they do as well as targets, deadlines and tasks (and people who attach equal importance to each).

STAGE 4: IDEAL MODELS (“CONCEPTUAL MODELS” IN CLASSIC SSM)

If we were able to do it just as we wanted, what would all the steps look like? Like all models there are rules for constructing this ideal. In SSM there are two basic rules:

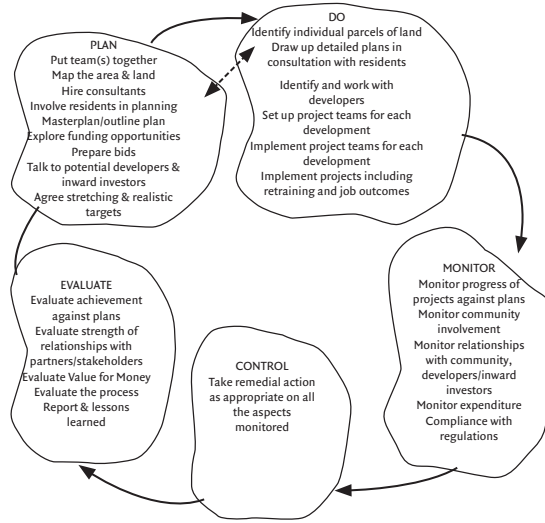
- The models must be solely and logically based on the root definition and *nothing else*
- The models must demonstrate the properties of a “system” (in Checkland’s case this includes “purpose”, “resources”, “continuity”, “monitoring”, “decision-taking processes”, “environment”, “boundary”, “interacting components”)

The activity headings in these models are based on Checkland’s recommended process of writing down the activities necessary to make the Transformation happen (Checkland and Scholes 1990). In one sense this is an expansion of the basic management cycle of “Plan, Do, Check, Act” that is widely used in quality management – with the added rigour of the two above rules.

In the Darnall example, shared thinking came up with the following ideal models, again shown here in abbreviated form, for the two root definitions above.

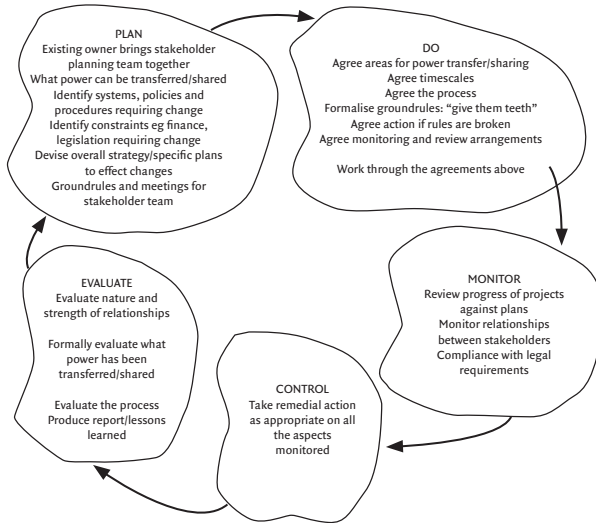
Figure 3: Ideal models

4a Relevant System for Root Definition 1: The Physical Regeneration of Darnall



4b Relevant System for Root Definition 2: Greater Community Control

[Note : This particular model here contributed to the approach taken under the a subsequent regeneration initiative]



I expanded these models in some detail, checking back against earlier stages and amending and improving as necessary, until I had mapped each necessary step in the ideal process. Finally, following the methodology, I checked that the models had the required systems properties (see above).

STAGE 5: COMPARISON WITH THE REAL WORLD

I then listed the differences between these two models with the “real world” as exhaustively as possible, with a couple of colleagues checking the list for any vital omissions. The same colleagues helped me choose which differences to focus on in the remaining stages, differences of “what” and “how”.

Checkland offers four suggested ways of how to do this: unstructured discussions, structured questioning (my preference and the most often used), scenario or dynamic modelling, and “trying to model the real world using the same structure as the conceptual model” (Williams 2005). The following table shows an extract of my comparison for the Darnall situation.

Ideal System (conceptual model)	Real Life	Difference
Evaluate the process	Project Management processes	No evaluation of overall processes
Information shared earlier with local residents about developments that will affect their lives	Information about specifics withheld for reasons of confidentiality	Skills gap between Council officers and some residents
		Mutual suspicion

I produced twenty or so pages of comparison, ensuring that I had covered most aspects. This is easy to do as a team exercise, can be quite good fun and certainly generates insights.

STAGE 6: OPTIONS FOR CHANGE

Checkland’s Stage 6 is developing culturally feasible and systemically desirable options. He suggests that at this stage, it can be useful to run through the model again using different CATWOE or modelling subsystems, also to carry out different systems based analyses. To ensure the feasibility and desirability of the options, Checkland further encourages you to analyse the situation’s ownership structure, the social & political conditions and power dynamics.

Personally, I do these in the stages from root definition onwards, so that I can generate a long list of options for change. I do this in order to keep both myself and stakeholders thinking freely, and to try to ensure that I have not missed a potentially significant option. In Darnall, I checked my list with colleagues who had more experience of the area, personalities and regeneration, and then reduced the original list to a manageable one of around twenty options, grouped under headings relating to the key tasks and issues, from which decision-makers could select their preferences. A brief extract is shown in the following table.

Possible change	Possible consequences if made/ not made	Change? (Y/N)
1. Reducing tension between planners and community a. Extend planning team to include legal, sales, community sections of local authority	<i>If made:</i> benefits could be greater harmonization of council activities, less confusion for community drawbacks could include the integrity of particular departments being compromised <i>If not made:</i> Existing confusion and anger on the part of residents continues and another mechanism must be found for managing the relationships	

Among other possible changes I listed were:

- Greater community involvement at all stages.
- Agreeing a workable definition of capacity-building for local use
- Developing the pool of skills available in the community to play an active and constructive part
- Holding inward investors to account for failure to deliver on local jobs for local people
- Different departments of the Council and regeneration organizations to work towards greater harmonization of aims and methods
- Those in authority who are charged with preparing plans and bids to include training and job targets at the earliest possible stage in the process

STAGE 7: AGENDA FOR CHANGE – SHARING AND USING THE FINDINGS

Checkland refers to this stage as “action to improve the situation”. In the real world, this is the point where the evaluator/analyst/consultant/facilitator usually hands over to the owners of the situation, who will then take responsibility for implementing any changes. Action may follow or it may not. However, the learning that takes place as a result of following the methodology may be considerable, at the individual, organisational and social level.

SOME CONSEQUENCES OF APPLYING SSM IN DARNELL

Since the SSM process was carried out towards the end of the physical regeneration programme in Darnall, little change happened at the time the formative evaluation was completed. Yet, for a small piece of work, the spin-offs from using SSM were considerable. Here are some examples of how certain recommended actions in the agenda for change were implemented:

When new European Community funding arrived, evidence about the successes and limitations of the previous regeneration programme was ready and waiting. There was significantly increased community participation in the Sheffield partnership. Darnall together with its neighbouring area of Tinsley for once gained a good share of available resources for capacity building.

The findings of this small exercise also contributed to the way in which European Community funding was evaluated in Sheffield. In the rest of South Yorkshire, the emphasis of the programme had been on capital investment, land reclamation

and major physical development. Sheffield's experience afforded consultants the opportunity to compare and contrast those approaches with alternatives.

Some final thoughts

Benefits of using SSM as an evaluation approach include :

1. It takes social, political, and power issues into account (which are in reality the driving forces behind many evaluations but may distort findings in the name of utility)
2. It is good for incorporating different perspectives on the success or otherwise of what is being evaluated –stakeholders have many different perspectives
3. It provides learning for future programmes and their evaluations (as here)
4. It is suitable for situations where there is considerable complexity accompanied by little clarity, with or without externally imposed targets
5. Sharing the process helps to facilitate clear thinking and choices
6. Allows for new and creative solutions to be discovered
7. Suitable for formative and summative, informal and formal evaluations.

In the context of the community sector, there are few disadvantages. The methodology is quick, logical and can be followed by an individual. It makes handling a surfeit of highly complex information relatively painless, incorporating statistical methods and data as appropriate, revisiting and reiterating stages to ensure robust findings. Voices that often go unheard in the sector and in evaluations (the voices of quiet dissent) can be given expression. There is greater transparency about the political processes at work.

Although Bowen and Shehata (2001) cite one of SSM's drawbacks as "heavy weight and time consuming process", the reality for the community sector is different. Much of our time is spent with individual clients and groups, which may be "heavyweight and time consuming" but is the lifeblood of what we do. Working through an application of the methodology can take up to two weeks' actual time – as distinct from elapsed time – in a sector that is always under pressure.

However, compared with the difficulties and time involved in raising funds to conduct formal independent evaluations, soft systems can be the most economic, efficient and effective way of carrying out a timely evaluation and learning lessons for subsequent actions and programmes.

If you are interested in using soft systems in evaluation, sources of further information include the user-friendly Open University material, chapter 6 of Checkland and Holwell (1998), or issues of the *Computing and Information Systems Journal*. In my experience, people generally prefer learning to use SSM by working through practical examples in real-life situations, as it is a technique that lends itself to learning by doing in order to keep the theory rooted in reality.

Further reading:

- Bowen, M and Shehata, S. 2001. SSM presentation notes presentation 29 November 2000.
<http://sern.ucalgary.ca/ffibowen/613/report/#prosAndCons>
- Checkland, P. 1981. *Systems Thinking, Systems Practice*. Chichester: Wiley.
- Checkland, P and Scholes, J. 1990. *Soft Systems Methodology In Action*. Chichester: Wiley.
- Checkland, P and Holwell, S. 1998. *Information, Systems And Information Systems: Making Sense of the Field*. Chichester: Wiley.
- Eden, C, Cropper S, and Ackermann F. 2005. *Getting Started with Cognitive Mapping*.
<http://www.banxia.com/depaper.html>
- http://www.open2.net/systems/practice/files/pet_chec.pdf. a succinct summary of the early development of soft systems methodology
- Patching, D. 1990. *Practical soft systems analysis*. Pearson Higher Education.
- Williams, B. 2005. *Soft Systems Methodology*. <http://users.actrix.co.nz/bobwill/ssm.pdf>

USING SYSTEMS CONCEPTS IN EVALUATION

Using Dialectic Soft Systems Methodology as an Ongoing Self-evaluation Process for a Singapore Railway Service Provider

Dr Boon Hou Tay
& Mr Bobby, Kee Pong Lim

At first reading you might be tempted to see this chapter as peripheral to evaluators' interests. Ostensibly it is about the design of a training program, not the evaluation of a training program. However, the chapter does four very interesting things. Firstly, it gets to the core of what a soft systems approach is trying to do – revealing in a rigorous rather than arbitrary or ideological way the multiple perspectives different stakeholders bring to a situation. Secondly, it explores the importance of using dialectic rather than dialogue as a means of gaining valuable insights and evaluative judgements from those perspectives. Evaluation has much to learn from both these features. Thirdly, it demonstrates the close links between the systems tradition and the action research tradition. Finally, the chapter shows how the original focus of soft systems methodology on large-scale program improvement can be used in much smaller scale circumstances.

Summary

This paper describes how to apply Dialectic Soft Systems Methodology (DSSM) as a self-evaluation process to assess the practice against a set of guidelines adopted by a Singapore railway service provider. This process ensures actual practices do produce the results that they intend on one hand, and to be able to review and refine the guidelines against missing gaps that may encounter during actual practices on the other hand. In this case these revised procedures are translated into a training system to be disseminated to the relevant parties on the ground as well as new trainees. It is the latter stage that turns this self-evaluation into an ongoing process by reviewing trainees' feedbacks and comments in each training session against the feasibility and desirability of the information gathered in the former stage. This ongoing self-evaluation process offers small-scale evaluation to be carried out by staff and management as part of their daily work activities. This process helps them to collect and use gathered data to answer their own questions concerning the quality and direction of their work. Further, this process can be applied to any workplace practice that is problem-focused and context-specific.

Introduction

Real life situations and problems for a railway system are complex and dynamic. In coping with them, we must understand the context within which they exist. One possible way is to conduct a self-evaluation on a set of existing practices (“the actual processes”) against a set of guidelines (“a set of ideals”) adopted by the

railway service provider. This self-evaluation process offers a view to incorporate different opinions of stakeholders and develop new understanding. It also pays attention to coalitions created especially for tasks that require the coordination among staff from different departments; to relate well with existing practices with what stakeholders know intuitively; and to provide an opportunity to express what each stakeholder know to others. This paper describes the use of Dialectic Soft Systems Methodology (DSSM) for such a self-evaluation process. DSSM is a small-scale evaluation conducted by staff and management as part of their daily activities. This process can assist them to collect and use gathered data to answer their own questions concerning the quality and direction of their work. Apart from improving practice, this process also strengthens a staff's timeless qualities such as confidence and capacity to think systemically. Further, this process can also be applied to any workplace practice that is problem-focused and context-specific.

This paper is organised into the following sections:

- Checkland's Soft Systems Methodology (SSM)
- A Different Description – Dialectic Soft Systems Methodology (DSSM)
- Dialectical Processes
- A Case Study for applying DSSM on a railway practice

Checkland's Soft Systems Methodology (SSM)

Peter Checkland and his colleagues developed Soft Systems Methodology (SSM) at Lancaster University in the 1970s using action research with an industry partner¹. This methodology was derived through collaboration with industry to address “soft” problems in social systems in which goals were often obscure as distinct from “hard systems” that were goal-directed. Hence the name soft systems methodology or SSM.

Most people use Checkland's seven-stage model as described in the work of Checkland (1981), Wilson (1984), Checkland and Scholes (1990), Checkland and Holwell (1998), Checkland (1999), Currie and Galliers (1999), Flood (1999), Dick (2000), Wilson (2001), Curtis and Cobham (2002), and Maani and Cavana (2002).

As described in Kate Attenborough's previous chapter the seven stages are:

1. The problem situation “unstructured”
2. The problem situation “expressed”
3. “Root definition” of relevant systems
4. Build “conceptual models”
5. Compare the “conceptual models” with the “real” world.
6. Think about feasible, desirable changes
7. Take action to improve the problem situation.

¹ See Checkland (1999), Dick (1999) and Sankaran, Tay and Cheah (2003).

SSM involves considering the problem situation in both the “real” world (Stages 1 and 2) and the “model” world where systems thinking is applied to develop root definitions to clarify the real problem and conceptual models are developed to look at ideal solutions (Stages 3 and 4). The “ideal” models are then compared to the actual situation. Differences between the models and reality become the basis for planning changes (Stages 5, 6 and 7).

A Different Description – Dialectic Soft Systems Methodology

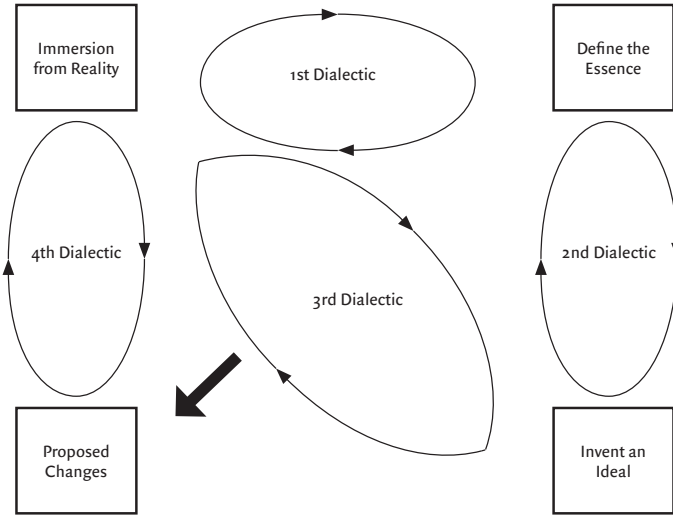
As highlighted by Flood (1999), Checkland (1999) and Tay (2003), Soft Systems Methodology is not a method that can be laid out in a set of stages to follow systematically. Checkland was fully aware of this difficulty when he formulated the “7-stage” model to act as a pedagogical tool to put forward Soft Systems Methodology Principles. Considerable effort was made by Checkland to explain the model as a continuous process of learning with which researchers begin anywhere and move in any direction even when it is explained within the limitations of linear prose.

The need for the “7-Stage” model to be understood as a learning cycle has prompted Dick (1993) to think of soft systems methodology as progressing through four dialectics.

However, it is important for readers to note that Dialectic Soft Systems Methodology (DSSM) is not a new form of Checkland’s Soft Systems Methodology. It is the same process as the “7-stage” description except it is presented from a different perspective. As described in the work of Dick (1993), Dick and Swepson (1994), Dick (2000), Sankaran, Tay and Cheah (2003), Sankaran and Tay (2003), Tay (2003), and Tay and Lim (2004), this approach makes explicit the inherent cyclic nature of Checkland’s seven stages and the use of dialectic comparisons. It progresses through four dialectics (see Figure 1).

- 1st dialectic – Between immersion (rich picture) and essence (root definition), where researchers try and experience the problem situation as fully as possible and then stand back and define its essential features (ie between Stages 1+2 and 3 of Checkland’s model).
- 2nd dialectic – Between the essence (root definitions) and the ideal (conceptual model) where the researcher try to find an ideal way to achieve the same transformation of inputs into outputs (ie between Stages 3 and 4).
- 3rd dialectic – Between ideals and reality where researchers think about improvement to the ideals or the actual situation (ie between Stages 4 and 5).
- 4th dialectic – Between plans and implementation where the plans are implemented and differences between plans and reality can be monitored through which further improvements can be carried out (ie between Stages 5 and 6+7 and back to Stage 1 in an action research cycle).

Figure 1: Dick's Dialectic version of Checkland's Soft Systems Methodology.



Dialectical Processes

At the heart of DSSM is the use of dialectical processes. As pointed out by Dick (1999), it helps in understanding *dialectical* processes to distinguish them from two other forms of processes in general use, namely, *adversarial* processes and *consensual* processes.

In adversarial processes, according to Dick (1997, 1999), it is common for one or two points of view to be accepted in its entirety. This approach therefore favours adversaries who are willing to put forward any information that supports their own view. Unfortunately, it also favours those who tell plausible lies, or at least selective truths. Adversarial processes thus embody the features of what Argyris and Schön label as “Model 1” and according to Dick (1999) they may be summarised as

- define goals unilaterally
- maximise own outcomes
- minimise the expression of negative feelings, and
- maximise rationality, minimise emotionality.

Consensus processes, on the other hand, begin by identifying the information on which people are agreed. The agreed information is then used as the springboard into more detailed planing or decision-making. The disadvantage is that there may not be enough agreement – or if there is, that people are reluctant to threaten it with the truth.

Consensus processes differ from Arygris' Model 1 in that goals are jointly agreed, and joint outcomes are pursued. The phrase "win/win" is often used to capture the essence of this. Negative feelings, however, are still minimised. Emotionality is tolerated, but only if positive.

As pointed by Dick (1997, 1999), it appears that both adversarial and consensual processes have something to offer. Adversarial processes may encourage a willingness to speak out, though they lead to a very selective use of information. They depend on competitive motivations. It is often an advantage if two people can be persuaded to speak willingly about disagreements. On the other hand, consensual processes develop a striving after agreement. They depend more upon our cooperative motives. This too, can be useful. However, it can be at the cost of discrepant information being denied.

According to Dick (1997, 1999, 2000), and Dick and Swepson (1994), dialectic processes try to combine the advantages and minimise the disadvantages. To do this they manage both sets of motivations, to compete and to cooperate. They encourage valid information while managing process to avoid its negative consequences. In this respect they share some common features with what has been called empathetic assertion ("the power of identifying oneself mentally with a person"). It is often a combination of assertive skills and listening skills that can provide the best combination, just as it is a combination of adversarial and consensual processes that are often effective in collecting valuable information.

In summary,

- Adversarial processes operate by choosing one of the competitive views.
- Consensual processes operate by identifying agreements.
- Dialectical processes use disagreements to generate agreement.

Dick (1999) describes dialectical processes as equivalent to Arygris' Model 2, where the followings are valued:

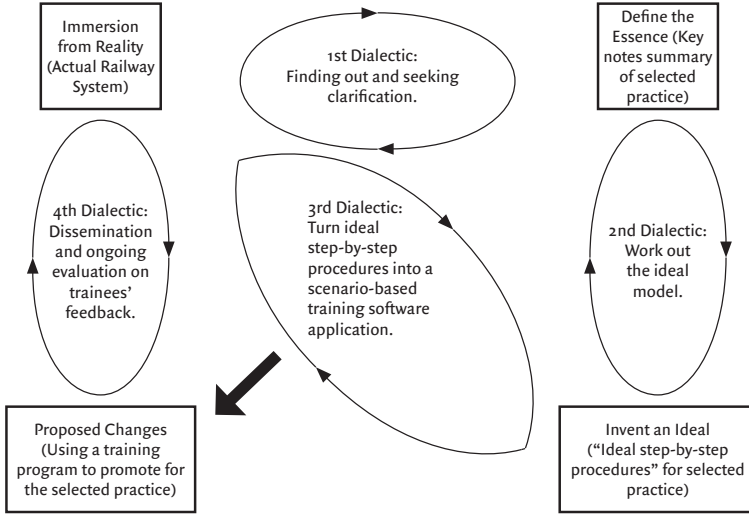
- valid information
- free choice and
- internal commitment and constant testing.

During the dialectical process, participants learn to balance understanding and judging, describing and evaluating, and inquiring and advocating. Out of the dialectic between opposing views, greater understanding emerges.

A case study on applying DSSM on a railway practice

This section provides a concise and comprehensive description on how to apply the Dialectic Soft Systems methodology ("Dialectic SSM") for a chosen railway practice. Figure 2 depicts the overall DSSM concept adopted for that selected railway practice.

Figure 2: The adopted Dialectic Soft Systems Methodology.



BRIEF DESCRIPTION OF THE SELECTED RAILWAY PRACTICE

The selected railway practice for this paper is called the “12-Car Push-out” operation. It is adopted when a defective 6-car train breaks down along the tunnel between two adjacent stations. It involves the use of an assisting 6-car train to couple and push the defective train to the next nearest station in order to detain all passengers onboard the defective train. The phrase “12-Car” is coined from the fact a 12-car train is formed when the two trains are coupled together. In this selected railway practice, DSSM is used to evaluate the set of working instructions to be carried out by a driver should such incidence occur.

FORMING THE WORKING TEAM

The working team for this comprised three members. The team was formed using the three conditions described in the work of Dick (1999) as the criteria for DSSM.

- Condition 1: That, among the team members, at least one member has access to the whole body of relevant information.
 In our team, the first member was a trainer who was a “subject matter” expert herself. She has access to the whole body of relevant information and is able to discuss with experts from other departments when necessary. Thus, our first member is said to be the “most” representative for the selected railway practice in our team.

- **Condition 2:** That there is preferably some overlap between the knowledge bases of different members, so that there is some check on the information each provides.
In our team, an engineer (who has experience in a similar field) takes up the role of second member. Unlike the first member, the second member was new to the specific issue. Therefore, she is said to be a “very” representative of the practice under investigation in our team.
- **Condition 3:** That they can communicate their probably-specialised knowledge to members from different specializations.
The third member fills up this role. In this case, she was an Information Technology specialist interested on taking the insights developed to develop a training program that enables the insights to be disseminated. As she has the least direct interest in the situation concerned, she is said to be the least representative for the selected railway practice.

In this case study, the first member is the owner of the results and findings derived from DSSM. The second member is a project leader, a facilitator and an evaluator who oversees, facilitates and evaluates during the entire DSSM. The third member is responsible for turning the agreed steps into a training package. As pointed out by Dick (1999), it is important to note the fact that dialectic focuses on information that is discrepant. It seeks to use this to improve the final decision, for which it also draws on a motive towards cooperation and agreement. A premium is placed on judgments which people hold with some confidence, and which tend to disconfirm the emerging consensus. This is best achieved when team members of different views are selected.

1ST DIALECTIC: FINDING OUT AND SEEKING CLARIFICATION

It is important to recall that both the second and third members are new to the subject. Thus, the primary objective of this dialectic is to enable second and third members to learn as much as possible from the first member.

Here is a summary description of the process.

- **Introduction**
The facilitator (second member) explained the need to know the required steps for the selected railway practice (ie 12-Car Push-out), and the desire for both honesty and respect for other people. The focus is to understand the steps needed as they are. Second and third members could pause the process by any clarification on the subject concerned. This initial stage is atheoretical and tries to avoid reinforcing existing mental models. Therefore the facilitator stressed that everyone should put aside their preconceptions or any act of forcing of data into theoretical moulds.

- **Facilitate the session**
An important part of this session was that very few questions are initially asked. The facilitator kept the first member explaining by using minimal encouragements – nods, non-committal grunts, “YES”, “hmmm”, “OK”, and the like. Any specific questions were asked only when the first member “dried up”.
- **Note taking**
Very brief notes were taken consisting of cue words only, preferably without looking at the notes. This gives an aid to memory without interfering too severely with rapport. In addition, taking digital photos of the written content on whiteboard, the background of station, the path leading to the trains, and the panels of switches and buttons to be activated. Like note taking, taking digital pictures offers a very rich source of information without disturbing the rapport severely.
- **Summary of key points**
Towards the end of this session, the facilitator asked for a summary of the key points. These are compared mentally to the facilitator’s own memory. Any probe questions for clarifications were asked.
- **Perform CATWOE**
The root definition for the “push-out” procedure was decided by first, second and third members. Checkland’s mnemonic CATWOE serves as a good checklist for ensuring that the important features of the procedures are included and this in turn enhances their understanding in the subject².
 - the Customers (“the passengers onboard”).....who are system beneficiaries
 - the Actors (“Customer Service Officer onboard”).....who transform inputs to outputs
 - the Transformation (“steps to recover a defective train”).....from inputs into outputs
 - the Weltanschauung (“Visibility to Station Managers and Traffic Controller”).....the relevant world views
 - the Owner (“Railway top management”).....the persons with power of veto
 - the Environmental constraints (“high power lines, darkness in tunnel”).... that need to be considered

In this case study, CATWOE made second and third members recognise the need to consider the role of station manager and traffic controller in the selected railway practice.

2 See the chapters by Ken Meter and Kate Attenborough for a fuller description of the CATWOE in SSM.

- **Individual work**
Immediately after the session, second and third members each prepared a one page summary of the results of the session.
- **Compare notes**
Second and third members met to compare notes. Each reported on the results of the session. After a comparison of overlaps and disagreements, probe questions were devised. As highlighted by Dick (1999), probe questions provide the dialectical mechanism. They help to cope with the amount of information generated. To winnow out the critical information from the very large amounts of data collected. Overlaps and disagreements form the data for analysis. It works like this. If two mention a topic, then either they agree or disagree. If they agree, they develop a probe question to test the agreement. That is, they try to maximise the chance that they will learn when it isn't true. When there is a disagreement, the probe question seeks to explain it. The process and sample are checked for adequacy.

This dialectic was repeated with the first member until the second and third members understood the context needed.

2ND DIALECTIC (WORK OUT THE IDEAL MODEL)

This dialectic focuses on the consolidation and compilation of information gathered from the first dialectic. It was in this second dialectic that an ideal step-by-step procedure to be performed by the train driver controlling the "push-out" operation were derived.

Here is a summary description of this dialectic.

- **Individual work**
The second and third member each drafted out the ideal step-by-step procedures needed for the "push-out" procedure based on CATWOE. In addition, both members also incorporated safety procedures and notifications into their step-by-step procedures. The safety procedures ensure that all passengers and the driver can get out of the situation unharmed. Notifications provide "visibility" to the traffic controller in order for her to make the necessary arrangements and decisions to avoid collisions with other trains running on the same route. Getting each member to draft the procedures individually favours individual expression and imagination (or individual's ideal model) on the subject and this in term adds richness to the subject's content. It further sets up another dialectic to drive insights.

- Compare drafts
Second and third members met to compare their drafts. As in first dialectic, after a comparison of overlaps and disagreements, probe questions were devised to test agreement and to seek explanation for disagreement. Both of them then integrated their draft and findings from the probe questions into a single draft.
- Verify with first member
The integrated draft was verified with the first member – who until then had taken no part in the discussions (ie another source of dialectic). Refinement and missing gaps were incorporated into a further draft.

The expectation from this dialectic is a comprehensive “ideal step-by-step set of procedures” for the “push-out” procedure.

3RD DIALECTIC (TURN “IDEAL STEP-BY-STEP PROCEDURES” INTO A TRAINING PROGRAM)

This dialectic focuses on turning the “ideal step-by-step procedures” into real-life practice via a training program. It involves comparing the “ideal” with back with “real-life”:

- Acquire details and information
Re-visit the project’s site; review the “ideal step-by-step procedures” against the physical environment such as the station, the tunnel, the train and the staff involved. Refine and fill up missing information in the ideal procedures where necessary. In addition, to capture the relevant multimedia information such as the status of display panels, the exact position of switches, knobs and train doors and the value of relevant system parameters that have to be taken into account for the selected railway practice, the sound of train during moving and video clips for route to be taken by the driver of the assisting train in the “push-out” procedure.
- Validate the training program with staff from other departments
Apart from reviewing with first member, it was fruitful to engage other staff such as a Traffic Controller and a Station Manager to validate the completed training program. Again, we sought to identify overlaps and disagreements from these new audiences. We devised probe questions to test agreement and to seek explanation for disagreement.

From this dialectic emerged a completed training program that satisfied both the “ideal” and “real-life”.

4TH DIALECTIC (DISSEMINATION AND ONGOING EVALUATION)

This dialectic focused on promoting the proposed changes (ie “ideal step-by-step procedures”) to new trainees using the training program. Apart from completing the SSM cycle it also importantly connected Step 7 and Step 1 in Checkland’s original model – and thus promoted continuous cycles of SSM akin to an action research approach.

In this dialectic, each batch of new trainees was exposed to the completed training program. Trainees were encouraged to document their feedbacks in a given feedback form during and after their training session. If a trainee identified a critical gap, the entire DSSM was repeated to address this new gap in the “push-out” procedure.

Conclusion

This paper highlights the use of Dialectic Soft Systems Methodology (DSSM) as an ongoing self-evaluation for helping an evaluator or a group of evaluators acquire the key knowledge about a situation; how to evaluate it against a set of guidelines; how to refine that assessment; and how to share it with others.

As pointed out by Tay and Lim (2004), real life situations and problems for a railway system are complex and dynamic. The future is essentially unknowable in complex systems. We cannot predict higher order abilities such as capability any more than we can predict creativity. We can obtain the skills and knowledge to be a painter but a “work of art” is another issue. The challenge for project success is to implement processes (such as the use of DSSM as an ongoing self-evaluation process within the organization) designed to create optimum conditions for making decisions in the face of new, unforeseen problems. This can be described as fitness of purpose rather than fitness for purpose in the work of Stephenson (1994) and Hase and Tay (2004). Documenting the process so we can share and learn from it might be the best we can hope for.

Another challenge for performing self-evaluation in workplaces is to not be frightened of conflict and ambiguity but see these states as an opportunity for learning. Perhaps even the creation of instability provides the atmosphere for learning to occur. When we are confused and anxious we can ask the questions that lead to deeper learning.

References

- Checkland, P. 1981. *Systems Thinking, Systems Practice*. Chichester: John Wiley & Sons, Ltd.
- Checkland, P and Scholes, J. 1990. *Soft Systems Methodology in Action*. Chichester: John Wiley.
- Checkland, P and Holwell, S. 1998. *Information, Systems and Information Systems – making sense of the field*. Chichester: John Wiley.
- Checkland, P. 1999. *Systems Thinking, Systems Practice*. Includes a 30-year retrospective. Chichester: John Wiley.
- Currie and Galliers 1999. *Rethinking Management Information Systems*. Oxford: Oxford University Press.
- Curtis, G and Cobham, D. 2002. Expert systems and knowledge bases. *Business Information Systems: Analysis, Design and Practice*, Fourth ed. First published in 1989, second published in 1995 and third published in 1998. Harlow: Pearson Education Limited. 580–625.
- Dick, B. 1993. *You want to do an action research thesis?* [On line]. Available at <http://www.scu.edu.au/schools/gcm/ar/art/arthesis.html>
- Dick, B and Swepson, P. 1994. *Appropriate validity and its attainment within action research: an illustration using soft systems methodology* [On line]. Available at <http://www.scu.edu.au/schools/gcm/ar/arp/sofsys2.html>
- Dick, B. 1997. *Dialectical processes* [On line]. Available at <http://www.scu.edu.au/schools/gcm/ar/arp/dialectic.html>
- Dick, B. 1999. *Rigour Without Numbers. The potential of dialectic processes as qualitative research tools*. Queensland, Australia: Interchange.
- Dick, B. 2000. *Soft systems methodology. Session 13 of Areol – action research and evaluation on line*. [Online] Available at <http://www.scu.edu.au/schools/gcm/ar/areol/areol-session13.html>.
- Flood, R L. 1999. *Rethinking The Fifth Discipline: Learning within the unknowable*. London: Routledge.
- Hase, S and Tay, B H. 2004. *Capability for Complex Systems: Beyond Competence. Proceedings, Systems Engineering/ Test and Evaluation Conference 2004 in Adelaide, Australia*.
- Maani, K E and Cavana, R Y. 2002. *Systems Thinking and Modelling – Understanding Change and Complexity*. Auckland: Pearson Education New Zealand Limited.
- Sankaran, S, Tay, B H, and Cheah, Y S. 2003. *Application of a Dialectical Model of Soft Systems Methodology to Conduct Action Research. Proceedings, ALARPM/SCIAR Conference 2003 in Queensland, Australia*.
- Sankaran, S and Tay, B H. 2003. *Action Research Models in Business Research. Proceedings, 9th ANZSYS Conference 2003 in Melbourne, Australia*.
- Stephenson, J. 1994. *Capability and competence: are they the same and does it matter? Capability, 1 (1): 3–4*.
- Tay, B H. 2003. *Using Action Research to develop a Social Technical Diagnostic Expert System for an Industrial Environment*. Ph.D. Dissertation, Graduate Research College, Southern Cross University, Australia.
- Tay, B H and Lim, K P. 2004. *A Scenario-based Training System for a Railway Service Provider in Singapore. Proceedings, Systems Engineering/ Test and Evaluation Conference 2004 in Adelaide, Australia*.
- Wilson, B 1984. *Systems: Concepts Methodologies and Applications*. Chichester: John Wiley & Sons.
- Wilson, B 2001. *Soft Systems Methodology: Conceptual Model Building and its Contribution*. Chichester: John Wiley.

Evaluation Based on Critical Systems Heuristics

Martin Reynolds

Martin's chapter provides exceptionally important insights into a key but largely unexplored area of evaluation. How is "value" or "worth" decided and what are the consequences of that for the design, data collection, analysis and reporting of evaluations? The particular approach Martin describes, Critical Systems Heuristics, has been widely used in the systems field and is deeply evaluative. Arguably it is one of the most powerful evaluation frameworks yet developed, yet ironically it is almost unknown in the mainstream evaluation field. Deceptively simple (12 questions asked from two orientations) the complexity of what it does reveals itself layer by layer. That is why you will probably need to read Martin's chapter several times, not because it is difficult, but because each reading will reveal the next layer of this remarkable approach.

1 Introduction

Critical systems heuristics (CSH) is a toolbox of 12 basic questions embodying a set of principles for evaluating the (i) built-in values, (ii) power structure and (iii) knowledge-base, associated with a specified focus (or "system") of interest. At the same time CSH evaluates (iv) the moral basis on which a system operates as considered from the perspective of both beneficiaries and victims of the system. Whilst scientific data, including statistics, might be usefully channelled to support the output of a CSH evaluation, the overall evaluation is a qualitative exercise. CSH is nested in a constructivist tradition of systems practice primarily aimed towards collaborative improvement of, and developing responsibility over, the situation in which a system is embedded.

As suggested by the term "heuristic", CSH is a learning device. CSH follows a learning cycle which involves iterating between, first, making sense of a situation through identifying relevant boundaries a system of interest (cultivating holistic awareness), followed by a process of engaging with contrasting judgements (appreciating and developing perspectives) around such systems, and finally taking appropriate (responsible) action for improving the situation to which a system is serving.

CSH draws on the substantive work and philosophy of C. West Churchman, a systems engineer who, along with Russell Ackoff during the 1950s and 1960s, helped to define Operations Research in North America. Churchman pioneered developments in the 1970s of what is now known as "soft" and "critical" systemic thinking and practice in the domain of social or human activity systems. Churchman died in 2004. His legacy rests with signalling the importance of being alert to value-laden boundary judgements when designing or evaluating human activities. Boundaries are what we socially construct in designing and evaluating any human activity. Boundaries are made more explicit in moving perceptions from a situation of interest (for example, a mess, problem, an issue, or a more formal

entity such as an institution of some kind) towards a system of interest. A system of interest is a conceptual representation of boundary judgements associated with a situation of interest or concern relating to human activity. The efficiency and effectiveness of any system of interest depends on the actual boundary judgements associated with that system of interest. Hence, evaluating a system requires identifying the boundary judgements. So what are these boundaries and how might we recognise them in situations that we might wish to evaluate?

Churchman first identified 9 boundaries (conditions or categories, as he called them) associated with any human activity system of interest in his book *The Design of Inquiring Systems* (Churchman 1971). Human activity systems are first characterised as being purposeful. The primary boundary of any human activity system, as Churchman sees it, is defined by *purpose*. The other 8 boundaries relate therefore to a system's purpose. Churchman's work is characterised by a continual ethical commitment to the overarching purpose of improved human well-being. He later extended the 9 categories to 12 categories in a book provocatively entitled *The Systems Approach and Its Enemies* (Churchman 1979). This work significantly takes into account 3 extra factors (associated with "enemies") that lie outside the actual system of interest but which can be affected by, and thereby have an effect on, the performance of the system. Churchman grouped the 12 categories into four sets of three, each associated with a particular group of people relevant to the system (what I call "stakeholder groups").

In the early 1980s a doctorate student of Churchman from Switzerland, Werner Ulrich, drew up his own version of Churchman's original twelve boundary categories, and named the four groups in terms of four sources of influence on a system of interest – (i) motivation, (ii) control, (iii) expertise, and (iv) legitimacy (Ulrich 1983, pp 240–258). Each category stands for a boundary judgement that needs to be questioned both normatively (what "ought" ...) and descriptively (what "is" ...). As explicit questions, these were first published in Ulrich (1987) and later slightly reformulated in Ulrich (2000). Professor Ulrich worked with CSH in Switzerland as a public health and social welfare policy analyst and program evaluator.¹

The twelve sets of boundary questions represent the basic toolbox of CSH. This chapter provides some guidelines on how to operationalise the principles of CSH for purposes of evaluation, based on a retrospective examination of CSH in use. I will briefly describe the toolbox, along with suggestions on when to use it and the benefits of its use. I then present a brief analysis of how I applied CSH in a particular context to help clarify some of the methodological issues associated with using CSH and the kind of results that might be produced. I finish with some practical tips for the practitioner in developing skills in using CSH.

1 In 2005 Ulrich secured a three-year Visiting Professor appointment with the Systems Department at The Open University, and at the time of writing is working with the author on teaching and research projects.

2. The toolbox

2.1 CSH QUESTIONS

My own adaptation of the 12 boundary setting questions are summarised in Table 1. The key developments from Ulrich and Churchman relate to categories 8 and 9 in which I attempt to tease out more precisely the dimensions of “expertise”; a concern drawn from my own particular interest and experience in evaluating expert support (Reynolds 2003)

Table 1: Critical Systems Heuristic Questions for Evaluation

(adapted from Ulrich, 2000)

Sources of motivation	
1	Beneficiary (“client”): who ought to be /is the client or beneficiary of the service or system (S) to be evaluated?
2	Purpose : what ought to be /is the purpose of S?
3	Measure of success : what ought to be /is S’s measure of success (or improvement)?
Sources of control	
4	Decision maker : who ought to be /is the decision maker (in command of resources necessary to enable S)?
5	Resources : what components of S ought to be /are controlled by the decision maker (eg financial, physical, natural, human resources as well as social capital)?
6	Decision environment : what conditions ought to be /are part of S’s environment, ie not controlled by S’s decision maker and therefore acting as possible constraints?
Sources of expertise	
7	Expert (or designer) : who ought to be /is involved as providing expert support for S, ie providing some assurance or “guarantee” that the system can succeed?
8	Expertise (guarantor attributes) : what kind of formal and informal expertise or relevant knowledge ought to be /is part of the design of S and what ought to be /is providing “competence” or guarantor attributes of success for S (eg relevant technical or disciplinary support, consensus amongst professional experts, experience and intuition of those involved, stakeholder participation, political support...)?
9	False Guarantee : What ought to be / are false guarantor attributes of success; that is, possibly misleading forms of expertise that might generate a bogus or artificial sense of guarantee or validity (eg (i) superficial multidisciplinary input, and/or (ii) sole fixation on scientific data, statistics, or processes of deliberation and “consensual” populist viewpoints, and/or (iii) tokenistic, superficial claims to ideas of “empowerment”, “social responsibility” etc...)
Sources of legitimacy	
10	Witnesses : who ought to be /is representing the interests of those affected by but not involved with S, including those stakeholders who cannot speak for themselves (eg the handicapped, future generations and non-human nature)?
11	Emancipation : to what degree and in what way ought /are the interests of the affected free from the effects of S?
12	Worldview : what ought to be /is the worldview underlying the creation or maintenance of S? ie what visions or underlying meanings of “improvement” ought to be /are considered, and how ought they be /how are they reconciled?

Two features of Table 1 require elaboration.

1. The 3 questions associated with each source of influence address parallel issues: the first question (1, 4, 7, and 10) address issues of *social role*; the second question (2, 5, 8, and 11) address issues of *role-specific concerns*; and the third question (3, 6, 9, and 12) relates to *key problems* associated with roles and role-specific concerns. In more contemporary language, I would associate these terms with “stakeholders”, “stakes”, and “stakeholdings” respectively.
2. Each question is asked in two modes, thereby generating 24 questions in total. In CSH all questions need to be asked in a normative, ideal mode (ie what “ought” to be...) as well as in the descriptive mode (what “is” the situation...). Contrasting the two modes provides the source of critique necessary to make an evaluation.

These two features are represented in Figure 1 which might be used as a basic template for a CSH based evaluation. (Later, in Table 2, I will provide a set of “ought” mode responses to such questions in relation to a context of application.)

		Social Roles stakeholders	Role-specific concerns stakes	Key Problems stakeholdings
Sources of motivation		1 Beneficiary/ client	2 Purpose	3 Measure of success
	'is'			
	'ought'			
	critique 'is' against 'ought'			
Sources of control		4 Decision-maker	5 Resources	6 Decision environment
	'is'			
	'ought'			
	critique 'is' against 'ought'			
Sources of knowledge		7 Expert	8 Expertise	9 False guarantee
	'is'			
	'ought'			
	critique 'is' against 'ought'			
Sources of legitimacy		10 Witness	11 Emancipation	12 Worldviews
	'is'			
	'ought'			
	critique 'is' against 'ought'			

An important third feature of the 12 CSH questions also requires elaboration.

3. The twelve categories can also be delineated between an association with those *involved* in the system (CSH questions 1–9 associated with sources of motivation, control and expertise) and those not involved but otherwise *affected by* the system (CSH questions 10–12 associated with sources of legitimacy). In Churchman’s original terminology, this mirrors the division between “a system” and its “enemies”.

Now that you have the abstract layout of CSH, some initial health-warnings might be appropriate. Firstly, the questions will gain meaning when they are actually used in *practice* in a particular situation or context of interest. Secondly, the precise wording of the questions may change with respect to both (i) different contexts of use and (ii) preferred vocabulary of the user.

To help you appreciate the meaning behind the CSH categories I provide some brief explanation in my use of the terms – system, environment, and boundary.

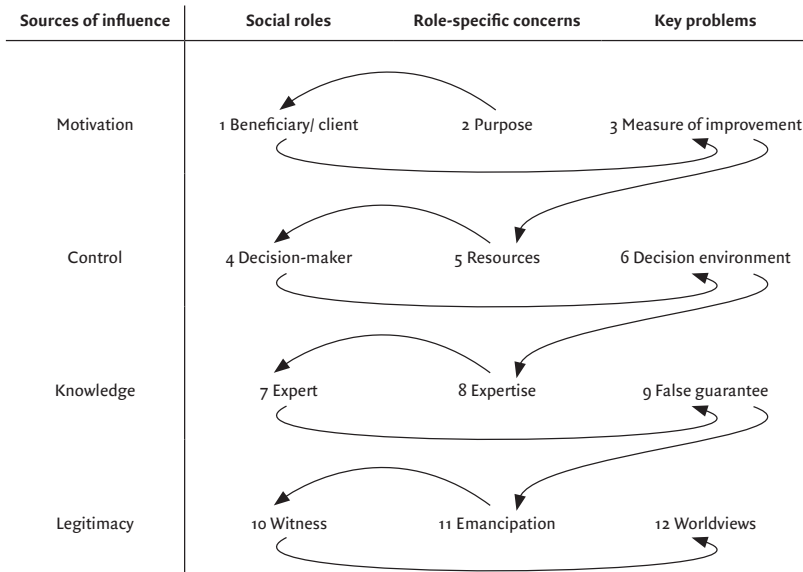
- i. A *system* (of interest) is understood as primarily a conceptual device (which may also approximate to a real-world entity). It represents a collection of interconnecting parts (or sub-systems) which function as a whole in order to *do* something (ie carry out a purpose). Systems might themselves be perceived as sub-systems of wider relevant systems, depending on the particular focus in interest. In CSH, “system” is considered as a heuristic device – one used for *learning* about and improving a particular situation, rather than making claim to *represent* a particular situation.
- ii. The *environment* of any system consist of factors outside the control of a specified system but which can nevertheless *affect* the system either (i) directly as when new government legislation may impact on, say, a system of accounting associated with a small enterprise, or (ii) indirectly as when welfare legislation is prompted by prior unethical actions of a large system such as the accounting procedures for a corporate entity, which in-turn affects the system’s activities.
- iii. A *boundary* exists (as with a system) in a conceptual sense between any specified system and its environment, or any sub-system and its immediate environment. Understood as human constructs, systems, subsystems, and their respective boundaries with the environment, are not absolute, but essentially open to judgement.

I will now describe in a little more detail the generic boundaries associated with the twelve CSH categories. For a more concise overview of the actual questions and their historic derivation from a rich tradition of practical philosophy, readers are directed to the original writings of both Churchman (particularly, 1979) and Ulrich (1983, 1988, and 2000).

A sequence of unfolding boundary judgements can be made attending first to issues of those involved in the system, beginning with sources of motivation,

and ending with attention to sources of legitimacy, as expressed through those affected by the system. Box 1 takes you through a short narrative of CSH evaluation addressing questions in the ideal “ought” mode. Figure 2 provides an illustration of the flow of unfolding associated with this narrative.

Figure 2: Unfolding sequence of CSH questions



The narrative in Box 1 illustrates a gradual unfolding shift in emphasis from core constituents of the system to features of its environment. In short, a full CSH evaluation enables stakeholders to step out of their immediate system of interest in order to see “the bigger picture”.

Box 1: An unfolding narrative of CSH evaluation in the “ought” mode (ideal mapping)

(The numbers in brackets refer to CSH categories 1–12).

Identifying first the *ideal* purpose (2) of the system of interest being evaluated in the “ought” mode, a CSH evaluation leads to an unfolding of key normative (ie ought mode) features. Stipulating the *intended* beneficiaries (1) and associated measures of success (3) associated with the purpose helps to make transparent the *value-basis* of the system.

Unfolding questions of motivation leads to questions regarding resources or components needed for success (5). Money and other forms of tangible capital assets might be complemented with less tangible factors such as access to social capital; that is, networks of influence where resources controlled by others might be aligned with the purpose being pursued by the decision makers (4). This prompts questions as to what relevant factors having an important potential impact on the system ought to lie outside the control of the system’s decision maker(s); that is, be a part of the system’s environment (6). For example, if a system initiated with good intention becomes malignant, corrupt or disabling for other more worthwhile systems of interest, are there factors in the environment that might ensure preventing the system to operate in such a disabling manner? Such questions help to make transparent the *power-basis* of the system.

One such set of factors requiring independence from the decision maker is “knowledge” or expertise. What are the necessary types and levels of competent knowledge and experiential know-how (8) required to ensure or, in-part, guarantee appropriate implementation? Who ought to provide such expertise (7)? How might such expert support prove to be deceptive or false (9), either by being incomplete (incompetent) or through assuming a dogmatic authority and complacency that does not allow for inevitable uncertainties (unforeseen events and unexpected consequences)? Such questions help to make transparent the *knowledge-basis* of the system.

Finally, given the inevitable bias regarding values (motivation), power (control) and even knowledge (expertise) associated with any purposeful system of interest, what is the legitimacy of such a system within wider spheres of human interests? In other words, if the system is looked at from a different, opposing viewpoint, in what ways might the activities be considered as coercive or malignant rather than emancipatory or benign (11)? Who (or what) hold such concerns, ie who are the “victims” of the system – and, importantly, what type of representation ought to be made on their behalf (10)? In other words, who may regard themselves capable of making representations on the victims’ behalf and on what basis would they make this claim? Finally, how might the underlying worldview associated with the system be reconciled with conflicting worldviews (12)? Where might representation of opposing views be expressed, and what action ought to happen as a result? Such questions help to make transparent the worldview or meaning underpinning the *moral-basis* of the system.

These last set of three questions are also crucial in exploring possible longer-term feedback or *systemic* effects of the situation being evaluated.

2.2 WHEN TO USE CSH

CSH can be used for straightforward outcomes-measurement evaluation of *purposive* systems of interest, where the purpose may be predefined and assumed unproblematic, and the emphasis is on evaluating the means (ie “summative” evaluation). But more significantly, CSH reaches beyond assumptions of consensus and evaluates the actual purpose(s) and implications of *purposeful* activity with relevant stakeholder groups (ie “formative” evaluation). As Ulrich explains: “purposiveness refers to the effectiveness and efficiency of *means* or tools, purposefulness to the critical awareness of self-reflective humans with regards to *ends* or purposes and their normative implications for the affected” (Ulrich, 1983, p.328). In other words, evaluating “means” ought not to be confused with evaluating purposes or “ends”. For example, counting the number of schools does not constitute evaluating regional or national education objectives! Whilst CSH has the capacity for evaluating the means towards ends, it also has capacity to re-explore the ends in changing situations.

It’s possible to identify four modes of CSH evaluation ²

1. Ideal mapping or *norms* evaluation: exploring only “ought” questions
2. Classic *summative* evaluation: contrasting “ought” with “is” questions to generate “critique”
3. Reframing or *formative* evaluation: allowing critique to inform new understandings and practices (new norms)
4. Challenging or *probing* evaluation: contrasting “ought” with “is” as a means of questioning dominant or hegemonic premises and judgements

The first three modes might also be considered as three sequential stages in a comprehensive formative CSH evaluation. The fourth mode is associated more with a cultivated and intuitive practice acquired through experience in using CSH. Most typically, CSH is used in the arena of evaluating *plans* or *planning processes* either as a *post-hoc*, summative evaluation, or as a more constituent *in-situ* formative evaluation. Both Churchman and Ulrich stress the importance of locating the planning process at specified levels in order to appreciate the selectivity or partiality of any purpose associated with planning. There are a number of different and sometimes confusing and conflicting terms used to describe these levels, but the important point is to appreciate the relative constraint imposed by each level, and to be able to locate your own system of interest within a particular level. The terms used below are different from Churchman and Ulrich, but appear in my view to resonate more with contemporary practice in planning. The three levels of planning are based on the principle of “vertical planning” originally suggested by Erich Jansch. In describing each level below, I supplement and paraphrase from descriptions made by Churchman (1979), Ulrich (1988) and Eden and Akermann (2001):

- *Operational planning* takes the purpose as given. The job is to define the

² Adapted from conversations with Werner Ulrich during 2005 and 2006.

exact means that will secure “improvement” in terms of the given purpose. Operational planning seeks *potential options* in implementing plans.

- *Objective planning* determines the purpose so as to secure improvement toward some overall vision of improvement, which is assumed to be given. Objective planning might be perceived in terms of developing *strategic directives and/or mission statements*.
- *Goal or ideal planning* can drop the feasible and the realistic and challenge the soundness of the visions implied by “realistic purposes”. Goal planning might be associated with the design of *vision statements*.

Churchman relates this last level of planning to the “ought” mode in unfolding his twelve categories³:

“In a sense it is ‘timeless’ planning, because its vision may be forever a vision. But through it the unfolding can begin in earnest, because its questions become more and more comprehensive: Why so many malnourished babies? Why the military buildup? Why so many dreary lives? Because, says objective-planning, there is no feasible way of changing these conditions. Such a reply – to the ideal-planner – is based on a given, not imagination. Since the main point of ideal-planning is the unfolding process, it is not even relevant to ask whether the ideal is ‘right’ in the same sense in which there is a right [operational] plan or a right objective plan” (Churchman 1979, p 83).

The three levels of planning can be helpful in positioning your chosen system of interest within a situation of interest, thereby allowing you to determine how purposive (summative) or purposeful (formative) the evaluation exercise might be.

2.3 WHY USE CSH?

There are three reasons for considering CSH for supporting an evaluation. These reflect broader principles associated with a tradition known as critical systems thinking (CST). CST draws particular influence from Churchman in the systems tradition as well as Jürgen Habermas in the social sciences domain.⁴ CST encompasses a wide range of methods from different disciplines, and whilst the focus here is on CSH, practitioners may well recognise other innovative evaluation techniques which clearly share affiliations with one or more of these three principles (see introductory chapter in this book).

1. *Cultivate holistic awareness*. CSH draws in a range of factors (Churchman calls this “sweeping in”) which may be considered. Conventional concerns relating to “measures of success” are linked here with important issues of “power” and “knowledge”, as well as “externalities”, including the influence of those affected by, but not involved with, the built-in design of

3 Churchman used “goal planning” in place of what I term “operational planning”.

4 For short readings associated with CST see Flood and Jackson (1991) and Flood and Romm (1996).

such measures. CSH prompts questions on who important stakeholders might be (“social roles”), and what their particular stakes (“role-specific concerns”) and stakeholdings (“key problems”) relate to. CSH can reveal important assumptions and premises which are often important to appreciating underlying “failure” in performance.

2. *Appreciate and develop perspectives.* The response to CSH questions leads to important reflection and triggers conversation around essential aspects of situational change. CSH questions can be used monologically, for reflective analysis, in changing your own perspective, and dialogically, for generating discussion amongst stakeholders around planning issues. Whereas conventional evaluation often limits stakeholders as subjects to evaluation, CSH allows for, and prompts the possibility of, change in stakeholdings amongst stakeholders.
3. *Nurture responsibility.* CSH enables questions to be raised regarding not only whether particular “objectives” are being achieved, but whether they are the right objectives to be sought after as viewed from the perspective of others, and what alternative objectives might be more appropriate. In short, CSH enshrines the notion of improved well-being as a trigger for the unfolding process. Of course, one person’s sense of improvement might be another person’s sense of deterioration. What is relevant is to have the question systematically raised amongst stakeholders as part of the process of evaluating, and for the CSH practitioner to take appropriate action. In this sense, CSH enables responsible practice.

3 The Technique: doing a CSH evaluation

Skill in CSH-based evaluation arises from practical use in the unfolding of CSH questions in different contexts. The technique of doing CSH varies between different practitioners with different interests and prior experiences of using CSH or similar techniques, and between different contexts of use. There is no prescribed methodology, but there is an adherence to the three principles outlined above.

The following process of conducting a CSH evaluation come from my own experience using the methodology in a range of contexts.

1. *Searching for “system”:* Identify a system of interest (SoI) for evaluating (ie the plan, task, project, programme, strategy, policy etc.) from the situation in which you are in. Name the SoI by addressing CSH question 2, assigning a higher-order (ie goal planning) ideal (“ought”) purpose to the entity being evaluated (ie “A System to...”). Note whether the evaluation is principally post-hoc summative or more process-oriented formative.
2. *Your role as evaluator?* Reflect and make a note on your own role as evaluator in the system of interest being evaluated. To what degree is your evaluation independent of the decision maker(s), or is there some possible compromise in the relationship which may inhibit independent appraisal? Might you be a mere human resource (category 4) enlisted to provide token reassurance for

some reason? Alternatively, do you consider yourself an “expert” associated with the system, engaged with providing *independent* expert support (category 7)? Or do you consider yourself more as a witness for the affected (category 10)? Possibly you may feel your role is a mixture of several of these. Record your first impressions to reflect on later in the evaluation.

3. *Level of planning?* For the SoI identified, attempt to locate where it fits within the three level hierarchy of planning: (i) operational, (ii) objective or (iii) goal/ ideal planning (see 2.2 above). This will give you some idea of whether the stated purpose is truly ideal (visionary), or circumscribed by some higher order purpose or vision.
4. *Initial CSH mapping:* Make a very rough first pass through of CSH questions in the “ought” mode. Use any initial reference material that you may have at hand including the terms of reference for your evaluation, and make notes of questions that you may wish to follow-up on during the process. When unfolding your system, address CSH questions in a systematic manner, beginning with questions of purpose (category 2). My own preferred sequence of “ought” mode questions for unfolding the SoI is 2nd then 1st then 3rd questions associated with each successive source of influence (see Figure 2 and Box 1). In other words, discussing the *key problems* associated with one stakeholder group generates questions relating to the *role concerns* associated with the successive stakeholder group which in turn triggers questions regarding who the stakeholders ought to be prompting questions regarding key problems associated with the stakeholding etc. After doing an initial “ought” mode mapping, undertake a similar, albeit very rough, “is” mode mapping from your first impressions of the situation and associated system.
5. *Identifying stakeholders:* From this first draft, suggest which individuals or agencies might best represent the main stakeholder groups associated with the SoI (ie categories 1, 4, 7, and 10) and provide examples of representative individuals or groups associated with each source of influence for possible interview. There will inevitably be some cross-over of interests associated with any one stakeholder identified. A government agency for example may claim to act in all four roles regarding a system for improving welfare development. The key point of this activity though is to get a general sense of which stakeholders are primarily concerned with particular role-specific concerns. So for example, a government agency in the context of a SoI for *health care* provision in America might primarily represent the “witness” category, whereas in the United Kingdom a similar agency might primarily represent the “decision maker” category. Depending on your capacity and resources available to you, the evaluation might be further undertaken through your own reflection on documented evidence monologically, as well as dialogically using conversations with stakeholders themselves. From your initial critique of “ought” and “is”, record questions/ issues that

- you might wish to follow-up with particular stakeholder representatives.
6. *Monological engagement*: Check further documentation and/or conversations with the commissioner of the evaluation in order to address the questions raised. Iterate between your template recordings and any further inquiry. Even where dialogical engagement is possible, there needs to be reflective time on the part of the evaluator in continually reassessing progress. Keeping a learning journal of the evaluation enables such reflective practice and contributes to more effective reporting (stage 8).
 7. *Dialogical engagement*: Interviewing stakeholders and inviting feedback from reports on interviews provides for both a more concise unfolding of the SoI from the perspective of the evaluator, but more importantly allows the opportunity for purposeful engagement amongst the stakeholders. Interview questionnaires can be designed for each of the key stakeholder groups identified. The questionnaire can be designed around CSH questions in two ways. In either way, it is important that (a) the terminology used in asking the questions is adapted for the particular context in which you are working, and (b) the questions are around issues regarding the purpose of the SoI in focus. Firstly, the questionnaire might be structured to systematically unfold a perspective of the SoI from each stakeholder group through adapting all twelve CSH questions in the same unfolding sequence as suggested in Figure 2. Alternatively, you might like to start your conversation with the “role-related concerns” associated with the particular stakeholder group that you are addressing. These concerns derive from your first monological unfolding of the SoI. If time and interest permits, you may like to expand the conversation to include viewpoints on role-related concerns of other stakeholders involved in and affected by the SoI. There is no one formula for designing questionnaires around CSH. Much depends both on the context of use and the experiences and intuition of the user. Some overarching guidelines are appropriate in order to foster trust and purposeful engagement: (i) brief your interviewee beforehand on your role and purpose in the evaluation process; (ii) provide some indication of (or possibly invite opinions on) the level of confidentiality involved with the conversations; (iii) indicate opportunities for future feedback; and (iv) use a semi-structured format, making explicit from outset (or even prior to meeting) a very rough picture of where you want to lead the conversation but without imposing constraints on issues that the respondent might wish to follow. Skill in using CSH questions is evidenced by not sticking rigidly to the format of questionnaire being used at the outset. Finally, whilst providing an atmosphere of ease in conversation, you need also to keep alive a level of tension implicit in the types of questions generated by CSH. My experience in using CSH questions amongst different professional groupings suggests that respondents generally enjoy being challenged.
 8. *Reporting*: Evaluation using CSH, whilst sometimes done in a summative

post-hoc context, is an essentially iterative learning process. A key task is to engage stakeholders in a continual reflective learning cycle around the system of interest in order to develop a sense of mutual development around purposeful collective activity, rather than developing fears associated with an “inspection”. Reporting will need to be undertaken in a clear narrative form. Simply presenting 12 sets of critiques will not make much sense. My advice in writing a narrative is that (i) your own role as evaluator is clearly registered (ie which views are yours and which views are assumed?); a useful device for this is to write in the first person singular (using terms like “I” and “in my view”); (ii) reference to a normative “ought” is clearly explained (and open to challenge); and (iii) crucially, you present your evaluation as an invite for further comment and deliberation.

4 Case Study: natural resource management

I now present a brief retrospective summary of an extensive evaluation exercise made during fieldwork in Botswana in the mid 1990s. The case study is actually an evaluation of three projects over a relatively long period of time (two years), with a substantial number of interviewees (78), many of whom (24) were interviewed on two separate fieldwork occasions. Whilst such an extensive and prolonged exercise is clearly not typical of the remit of a professional evaluator, the example provides a picture of the general features associated with a full CSH evaluation from which you may acquire some ideas for developing your own practice. My aim is to briefly illustrate the techniques employed rather than to detail the substantive outcomes. For further details on process and outcomes see Reynolds (1998). The notes are ordered in the same sequence of technique stages outlined in the previous section. The category numbers referred to are CSH categories 1–12.

4.1 IDENTIFYING A SYSTEM FROM A SITUATION OF INTEREST

Botswana is often cited as an African economic success story. Economic planning has been based principally on the trickle-down strategy of using revenue from a rich source of *non-renewable* diamonds to finance public sector expansion and improvements in rural infrastructure including provision of health, education, agriculture and communications. The impact of planning *renewable* natural resource-use is less impressive, as evidenced by persistent high levels of rural poverty amidst a diminishing and degrading stock of communal (as against privatised) natural resources.

Since the early 1990s, considerable attention has been given to promoting participatory planning in less-developed countries as a means of poverty alleviation and protection of the natural environment. In Botswana, participatory planning was being extensively piloted as a means of natural resource-use appraisal in rural areas during the 1990s with the support of donor agencies and the national government.

My situation of interest was the role of participatory planning in rural development. My system of interest (SoI) for evaluation was: *A system to enhance natural*

resource-use appraisal (NRUA) through participatory planning for assisting rural poverty alleviation and protection of the natural environment in Botswana.

The evaluation was intended to be more “formative” than “summative”, as my input became part of a wider ongoing appraisal of participatory planning in Botswana. In terms of another system of interest with a different purpose – relating to the completion of my doctorate thesis – the evaluation might be regarded as more summative.

4.2 ROLE OF EVALUATOR

I was not commissioned or paid for by any stakeholders associated with the system of interest, and so can claim a fair degree of independence. My own source of support derived from the UK Economic and Social Science Research Council which financed my fieldwork as part of a wider package of support for doctorate studies. The reports produced were written and presented to the stakeholder representatives without prior conditions. My own role then as an evaluator was closely associated with both categories 7 (“expert”) and 10 (“witness”) relating to the SoI described above.

4.3 LEVEL OF PLANNING

Three separate ongoing projects were chosen for evaluation successively representing three progressively wider domains of planning:

- i. Participatory Rural Appraisal (PRA) Pilot Project (operational planning)
- ii. Natural Resource Management Project (NRMP) (objective planning)
- iii. Botswana Range Inventory & Monitoring Project (BRIMP) (goal/ ideal planning)

Whilst occupying different levels of planning, each project shared important features: firstly, their prime objectives are social and environmental rather than economic; secondly, significant direct or indirect non-governmental sources of expertise (non-government organisations, private consultants and parastatals) – reinforced with donor support – were commissioned; and thirdly, each project promotes the use of “participatory techniques”.

In effect there are three systems of interest being evaluated. Each nested within a particular level of planning.

4.4 INITIAL CSH MAPPING

Before exploring individual projects, a sketch of the normative use of participatory planning for NRUA practice in Botswana was mapped out, based principally on background reading of the situation. Table 2 illustrates an ideal mapping exercise of NRUA in general using the 12 CSH categories in the sequence as described in Figure 2.

Table 2: Ideal mapping of participatory natural resource-use appraisal in Botswana

Sources of influence	Role	Role-specific concerns	Key problems
Motivation	Beneficiary Rural poor, future generations and non-human nature.	Purpose To improve natural resource use planning in addressing needs of the vulnerable	Measure of improvement Indices of i. rural poverty alleviation and ii. enhanced condition of natural resources
Control	Decision-maker Communal resource users	Resources Necessary components to enable NRUA; including i. project/ finance ii. human iii. social networks	Decision environment i. interest groups affected by NRUA ii. expertise not beholden to decision maker
Expertise	Expert Communal resource users informed by natural and social scientists and other sources of relevant knowledge/ experience	Expertise i. technical and experiential know-how & knowledge, including rural peoples' knowledge; ii. interdisciplinary and intersectoral facilitation skills iii. social & environmental responsibility	False guarantee Incompetent/ incomplete expertise and i. "scientism" (sole reliance on objective and statistical "fact"), ii. "managerialism" (sole reliance on facilitating communication), and iii. "populism" (allowing loudest collective voice as sole guarantor)
Legitimation	Witness Collective citizenry representing interests of all (including private sector) affected by NRUA, both local and global, and present and future generations.	Emancipation NRUA open to challenge from those adversely affected, including interests of private landowners and diamond industry competing for access to communal resources	Worldviews Manage conflict between interests of i. national economic growth/ privatisation & fencing policies, with ii. vulnerable rural livelihoods & nature

In the ideal world of purposeful human activity, the roles of beneficiary, decision maker, expert and witness are closely interrelated and at-one together. For natural resource-use appraisal, a system of self-organisation and appraisal amongst conscientious natural resource users might therefore be considered as the ideal situation. However, given that we do not live in an ideal world, a descriptive map was then required in order to identify who the stakeholders might actually be.

This initial ideal mapping provided a benchmark for developing further iterations of normative mapping at each level of planning, as well as providing the basis to critique "descriptive mapping" when evaluating each of the three projects.

4.5 IDENTIFYING STAKEHOLDERS

Four institutional types were identified as representing generic social roles of beneficiaries, decision-makers, experts, and witnesses associated with NRUA in Botswana. These are, respectively, government departments, donor agencies, consultants, and non-government organisations (NGOs) (see Table 3). Whilst impoverished natural resource users would clearly represent the ultimate “ideal” or *intended* beneficiaries (see Table 2), for the purpose of identifying *actual* stakeholders associated with each project, I wanted to address and interrogate the *immediate* beneficiaries of NRUA – the various government departments who would claim to be working on behalf of the rural poor. During the evaluation I kept in-check assumptions that (a) government would make appropriate representation of such stakeholders, and if not, (b) NGOs would ensure such representation.

Table 3: Stakeholder map associated with natural resource-use appraisal in Botswana

Institutional type	Primary role in NRUA projects
Government departments	<i>Beneficiary</i> : improved NRUA practice for better delivery on, and design of, government policy, on behalf of rural constituency.
Donor agency	<i>Decision-maker</i> : providing resources efficiently for effective NRUA practice
Consultancy (academic or private business)	<i>Expert (professional)</i> : ensuring impartial production of knowledge for sustainable and ethical natural resource use
Non-Government Organisations (NGO)	<i>Witness</i> : representing interests of impoverished natural resource users, future generations, and non-human nature

The generic roles assigned to these institutions are not mutually exclusive, but whilst there might be considerable overlap in roles and role-related concerns, it was useful to have this first mapping of stakeholders as a basis for starting a more detailed evaluation of NRUA associated with each project.

4.6 MONOLOGICAL ENGAGEMENT

Compared with professional practice in evaluation, my own engagement in this evaluation afforded me ample time for engaging with grey (ie non-published) literature material associated with the projects, and reflection on interview notes between interviews. In this time I kept three journals in relation to each of the projects being evaluated. The material provided an essential resource for writing up the interim reports for my respondents to feedback on.

4.7 DIALOGICAL ENGAGEMENT

Rather than systematically going through each of the 12 CSH questions for each interview in both the “ought” and “is” modes, each schedule for this evaluation was customised according to (i) the perceived stakeholder role (beneficiary, decision-maker etc), (ii) the particular level of planning/ project being focused upon (often, interviewees would have a stakeholding in several projects at the same

time, though it was important to record level-specific notes where appropriate), and (iii) information arising from prior interviews with other stakeholders and/or relevant grey literature.

After introducing the focus of evaluation in terms of participatory planning for NRUA, each schedule began with questions relating to what the stakeholder considered to be their main role, and their main concerns and key problems in fulfilling their role. Time and interest permitting, wider questions were then asked about relationships with other stakeholders, and an impression of what the roles, concerns and problems associated with these stakeholders might be in the context of NRUA. The initial ideal/ descriptive mapping provided prompts in developing the conversation throughout the interview. Conflicts between respondents associated with the same stakeholder group were recorded and used for further enquiry and/or included in the interim reports.

In recording feedback from such conversations, it was useful to continually update the impression of what “is” the situation with respect to each level of planning. In other words, the descriptive mapping is a continually evolving exercise during conversations and any associated reading of informal grey literature revealed and made available from such conversations. At the same time, critiques were emerging from the descriptive mapping. This is essentially a subjective exercise on the part of the evaluator. It was important to keep a record of the developing critique as this became the basis for reporting back. There is not the space here to look at any actual descriptive mapping associated with the projects, though Boxes 2–4 give some indication of the final critiques that emerged from the mapping exercise.

4.8 REPORTING

Reporting back on a CSH-based evaluation requires transparency as well as skill in translating findings and impressions in a mutually appreciated vocabulary and narrative. A key to successful evaluation is in eliciting recognition and critical appreciation and further engagement amongst the stakeholders involved. Along with fieldwork observation of participatory rural appraisal techniques in operation, and analysis of a substantial amount of grey material associated with each project and level of planning, inevitably this exercise generated a large amount of data and information to assimilate. Keeping a journal of the critique became a particularly important feature of this particular evaluation, along with the development of a series of three successive interim reports submitted back to the stakeholders which provided important feedback for further iterations. Three interim reports were submitted during the course of the two years.

Each report began with an explicit statement on (i) my own role and purpose with respect to the evaluation exercise; (ii) a disclaimer regarding any pretence to having made a “scientific” evaluation; and (iii) what I perceived were the main issues of the evaluation, couched in terms appreciated by the stakeholders (ie underlying values and purpose of the project, issues of relevant power and decision

making, relevant knowledge, and moral underpinnings).

All stakeholders were invited to comment on the interim reports either through written submission and/or verbal communication through either further private communication or special discussion sessions (one exclusive seminar for a government committee on rural development extension practice, and one public seminar were specially convened in Botswana for such feedback).

Boxes 2 to 4 provide very brief summaries of the final critique presented for each respective project. Each embellishes some descriptive mapping and specific critique of “role”, “role-specific concern” and “key problems” associated with each source of influence (ie as derived from template in Fig 1).

Box 2: Participatory Rural Appraisal (PRA) Pilot Project (operational planning)	
Motivation critique	Local government extension officers were immediate beneficiaries rewarded with facilitation skills to enable greater involvement of local people in extension work. But to what extent might <i>alleviating perceived rural social inertia</i> lead to poverty alleviation? The key measure of success for the project was centred on <i>high levels of participation</i> and generation of <i>self-help projects</i> . Perhaps instead, rural poor might benefit from better access to and control over resources rather than being subject to further consolidation of government extension practices.
Control critique	Under trajectory of (i) increased privatisation and fencing of communal land resulting in further alienation of natural resource, and (ii) reduced government assistance for local development projects, rural poor livelihoods are increasingly dependent on contracts with landowners and donor support for collective projects. Is there a risk that rural peoples knowledge loses its independence in becoming increasingly subject to government extension practice which itself is circumscribed by government central policy?
Expertise critique	To what extent might participation levels amongst rural poor in PRA exercises provide a guarantee for poverty alleviation? Might this guarantor attribute distract from large body of empirical data and experience suggesting significant correlation between rural poverty and land fencing policy since the mid 1970s?
Legitimacy critique	Dominant underpinning belief that benevolent government (through tradition of generous handouts and transfer of technology projects) has been responsible for generating rural social inertia, hence the need for government to step back and allow “development from within”. Danger of further marginalising rural poor through not appreciating perceived root cause relating to control and access to land.

Box 3: Natural Resource Management Project (NRMP) (objective planning)	
Motivation critique	Key beneficiaries of NRMP appear to be management staff of community based natural resource management (CBNRM) projects responsible for eliciting support/ resources from different line Ministries (eg Wildlife & Tourism, Agriculture, Water Affairs, Local Government). But to what extent might improved <i>multisectoral planning</i> address rural poverty and communal land degradation? Key measure of success is the number of CBNRM projects, primarily as indices of improved intersectoral collaboration. But do CBNRM projects (i) use or simply bypass line ministries? (ii) elicit collaboration with government or dependency on donors? and (iii) serve the <i>very poor</i> ?

Control critique	Have CBNRM “projects” become new currency for rural development? Whilst CBNRM might appear to be better grounded in local needs, is there greater levels of accountability in use of financial resources as compared with government extension programmes? Does short term funding support from donor agencies allow government to divert resource support away from local rural development?
Expertise critique	CBNRM management requires multidisciplinary expertise and skills in facilitation. But to what extent are participatory techniques involving rural participants a useful trigger for intersectoral collaboration and communication between traditional sector and disciplinary based experts? Rural peoples knowledge may be regarded as a useful check on professional judgements but how far is it appreciated as a potential driver for rural development initiatives?
Legitimacy critique	Dominant underpinning belief that appropriate expertise ought to drive rural development rather than traditional dependence on civil service sector-based bureaucratic functions which inevitably create the closed “silo” mentality. Possible conflict with local understandings of the need for greater autonomy and control over development amongst rural participants in conjunction with locally-elected government officials rather than donor-promoted project managers.

Box 4: Botswana Range Inventory & Monitoring Project (BRIMP) (goal planning)

Motivation critique	Immediate beneficiaries are <i>policy advisors</i> wishing to instil <i>longer-term coordinated planning</i> to address problems of previous piecemeal development in rural sector. BRIMP is housed in the Ministry of Agriculture, dominated by free market neo-liberal economic development planning and policies associated with fencing communal rangeland. So how likely is it that such coordinated planning might benefit rural poor? Do economic measures of success associated with gross national (agricultural) product equate with rural poverty alleviation and enhanced condition of natural environment?
Control critique	Commoditised resources provide the most appropriate means for economic planning. Thus fencing of communal land, privatising water supply, project-oriented development, and having rural participants on-tap for consultations during monitoring and evaluation efforts, might be considered as important measures of control; consolidating existing relations of economic power rather than empowering the rural communities. Are there risks of further disenfranchising rural communities through consolidating private ownership of land.
Expertise critique	Central guarantee for ensuring properly co-ordinated efforts is through <i>purposeful monitoring and evaluation</i> using econometric indices based on criteria of efficiency and effectiveness in terms of generating economic wealth from natural resources. Participatory techniques using rural peoples knowledge, is regarded as a means of ground-truthing or checking information arising from more technically oriented surveillance systems like remote sensing.
Legitimacy critique	Dominant belief that free-market determinism using econometric devices applied to natural resource-use provide most effective means for reducing poverty and protecting the natural environment. Needs reconciling with Tswana tradition in communal rangeland management, and primacy of democratic debate as a means of determining policy.

Two features of these critiques need further comment. First, it may appear that the critiques are very negative. Does critique necessarily imply something bad, unworthy or wrong? And if it does, how can CSH evaluation hope to instil

purposeful engagement amongst stakeholders? The summaries provided here do not give justice to the creative aspects of each project. Nevertheless, a central task of CSH evaluation, in my view, is to nurture an attitude of creative disruption. From a critical systems perspective critique does not equate to being negative, but rather provides a platform (arising from a dialectic between notions of “positive” and “negative”) for improving our understanding and practice associated with a situation of interest. The reporting phase in a CSH evaluation is crucial in keeping on board the active interest and involvement of stakeholders. In hindsight, my own reporting at the time, though generating significant interest and emotional engagement in terms of attendance at follow-up meetings and quality of feedback, did alienate some key stakeholders involved with the projects. Under different circumstances, possibly as a professional evaluator with more active involvement in the projects, I would have been in a better position to further correct any adverse effects through follow-up work with the affected stakeholders.

A second feature of the critiques is that they signal some important points of inter-relatedness. In producing interim reports, I ensured that all three projects were presented together in a single document, drawing out critical questions regarding the wider system of interest. This was particularly appreciated by project personnel in (i) grounding their own work within wider spheres of related activity, (ii) gaining familiarity with, and exploring potential synergies between, closely associated projects, and (iii) enabling a platform for important critical feedback between stakeholders thus generating new stakeholdings in *changing* systems of interest.

5 Summary: reflections on skills development

CSH is not a prescribed methodology. There is a wide variety of practice in the use of CSH questions. In some circumstances, not all the questions may need addressing. Descriptive mapping might be appropriate before, or as a trigger to, ideal mapping. Ulrich himself uses CSH in slightly different ways in evaluating two substantive case studies – economic planning in President Allende’s Chile, and health systems planning for Central Puget Sound in North America (Ulrich 1983).

The key to developing CSH skills rests with appreciating the systems principles embodied in the tool: (a) the idea of boundary critique, in being systemically aware (and generating systemic awareness) of, and making explicit, the boundary judgements implicit in any human activity; (b) appreciating your own role and values relating to a situation of evaluation and the need for nurturing critical conversation amongst stakeholders to develop, rather than merely protect, stakeholdings; and (c) using CSH evaluation to serve wider ethical interests of well-being, both social and ecological.

More specifically, I offer some practical tips in the use of CSH arising from personal experience.

- i. Practice at deploying CSH questions is the only way of developing skills and appreciating the interrogative power of the questions being asked.
- ii. Practice using a system of interest relevant to you personally (eg a domestic

or work situation, activity, proposal in which there is some “purpose” attached).

- iii. Adapt the terminology to your own needs/ culture/ context of use, whilst retaining the essential meaning of the 12 categories.
- iv. Be prepared to encounter moments of discomfort or disruption in using CSH. Making values transparent is not a painless exercise, either for the evaluator or other stakeholders involved with evaluation.

This last advice suggests that skills in CSH evaluation requires associate skill-development in constructive engagement and constructive disruption!

Acknowledgements

Particular thanks go to my two reviewers, Mel Tremper and Leslie Goodyear, for their invaluable comments on an earlier draft.

Further reading

- Churchman, C W. 1971. *The Design of Inquiring Systems: basic concepts of systems and organizations*. New York, Basic Books.
- Churchman, C W. 1979. *The Systems Approach and its Enemies*. New York, Basic Books.
- Eden, C and Akermann, F. 2001. SODA – The Principles. *Rational Analysis for a Problematic World Revisited: problem structuring methods for complexity, uncertainty and conflict*. J Rosenhead and J Mingers (eds). Chichester: John Wiley. 21–42.
- Flood, R L and Jackson, M C (eds). 1991. *Critical Systems Thinking: Directed Readings*. Chichester, John Wiley.
- Flood R L and Romm, N R A (eds). 1996. *Critical Systems Thinking: Current Research and Practice*. New York: Plenum.
- Reynolds, M. 1998. “Unfolding’ Natural Resource Information Systems: fieldwork in Botswana. *Systemic Practice and Action Research* 11(2): 127–152.
- Reynolds, M. 2003. Evaluating Regional Sustainable Development submission to Workshop of the EU Thematic Network project REGIONET. Evaluation methods and tools for regional sustainable development. *Towards Systemic Evaluation: A Framework of Co-guarantor Attributes*. University of Manchester (UK). 11–13 June 2003.
- Ulrich, W. 1983. *Critical Heuristics of Social Planning: a new approach to practical philosophy*. Stuttgart (Chichester), Haupt (John Wiley – paperback version).
- Ulrich, W. 1987. Critical Heuristics of Social Systems Design. *Critical Systems Thinking: Directed Readings*. R L Flood and M C Jackson (eds). Chichester: John Wiley.
- Ulrich, W. 1988. Churchman’s “Process of Unfolding” – Its Significance for Policy Analysis and Evaluation. *Systems Practice* 1(4): 415–428.
- Ulrich W. 1996. *A Primer to Critical Systems Heuristics for Action Researchers*. Hull: University of Hull.
- Ulrich, W. 2000. Reflective Practice in the Civil Society: the contribution of critically systemic thinking. *Reflective Practice* 1(2): 247–268.

USING SYSTEMS CONCEPTS IN EVALUATION

Human Systems Dynamics: Complexity-based Approach to a Complex Evaluation

Glenda H Eoyang, Ph.D.

The area of complex adaptive systems (CAS) is perhaps one of the most misunderstood and misinterpreted in the entire systems canon. To some CAS is about complex computer simulations with little practical relevance, to some it is a quasi-mystic notion of unexplained “emergence”, and to some it promotes laissez faire approaches to social and economic change. Glenda steers a skilled, insightful and determinedly practical path between these turbulent ideas. She describes a task familiar to many evaluators (assessing an organizational change process) and shows how her blend of CAS and social science (Human Systems Dynamics) allows evaluators to understand what is going on. However, more importantly it challenges the laissez-faire idea that you cannot use CAS principles to suggest how to intervene and influence what might happen in the future – the applications of “simple rules”.

Introduction

Our challenge was to evaluate an organizational change process in a 3,000-person county social services department (Cope County ¹) and to recommend ways to move the change forward more effectively and efficiently. Several features of this project pointed us toward a systems-based approach:

- The business purpose for the organizational change was integration of human services, which involves collaboration and co-evolution among many programs that had traditionally worked independently.
- The organizational shift that was required to support integration of services merged six departments that were isolated from each other into a single mega-department.
- Multiple change initiatives were in process at the same time within the Department, and the organization needed a consistent way to evaluate each of the smaller projects and the context of the whole.

This kind of massive and increasing entanglement and the need to evaluate both small and large contexts in consistent ways signal to us the need for some systems-based approach for evaluation. Then the question emerges, “Which systems-based approach will be most effective?”

The diversity of systems approaches that are available to the evaluation professional today is well demonstrated by the variety of methods and methodologies represented in this volume. Distinguishing among their strengths

¹ Cope County Social Services Department is a fictional name for a real department in a real county. All other aspects of the case described here are true. Only the name has been changed in an effort to limit the effect that publication of this study might have on the internal and continually emerging dynamics in the real world of Cope County.

and weaknesses is a daunting task because definitions within each domain vary widely, and the explanations of the similarities and differences among them are even more numerous and confusing. When we select a method, we focus on distinctions that we have found most useful as practitioners looking to make sense of the wide range of systemic approaches available to us.

Conceptual Foundations

The assessment design for Cope County was based on systemic and conceptual assumptions drawn from two closely related fields of study: complex adaptive systems (CAS) and human systems dynamics (HSD). CAS is an interdisciplinary field in the physical and information sciences that explores emergent patterns of behavior. HSD applies principles from disciplines related to nonlinear dynamics (CAS is one) to see and influence the patterns of behavior as people work and play together in teams, organizations, and communities.

COMPLEX ADAPTIVE SYSTEMS (CAS)

Though it shares many characteristics with other systemic approaches, complex adaptive system views are different in some fundamental ways. A complex adaptive system is a *collection of semi-independent agents that have the freedom to act in unpredictable ways, and whose actions are interconnected such that they generate system-wide patterns* (Dooley 1996).

Emergent, system-wide patterns are said to “self-organize” in the system over time because changes in structure result from internal dynamics and interactions rather than external influences.

In the physical sciences, examples of self-organizing agents in complex adaptive systems (CAS) include bees that swarm, molecules that generate a Belousov-Zhabotinsky reaction, genes that shape a phenotype, and species that form a living ecosystem. In each of these cases, identifiable system-wide patterns result from the interactions between and among multiple agents and their clusters that form subsystems (Cowen et al 1994).

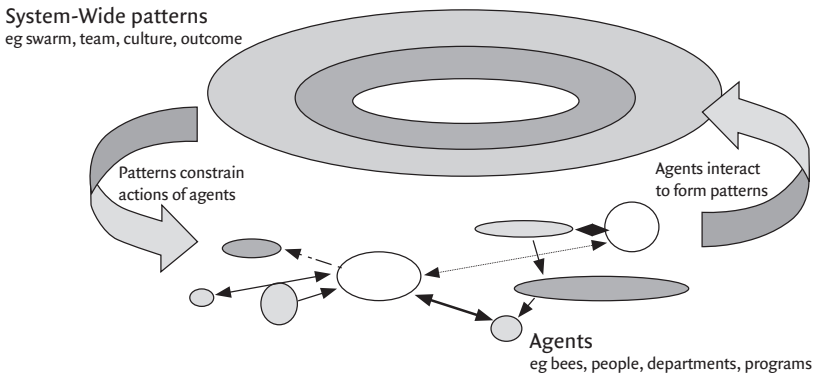
In human systems, analogous phenomena appear in the informal formation of teams, gangs, crowds, and cliques. The formation of emergent patterns also influences more formal systems, such as the cultures and structures of organizations, businesses, or governments. In each case, individuals or small groups interact in unpredictable ways, and system-wide patterns emerge over time.

The emergent nature of the CAS approach provides a flexible way to investigate and describe behavior, but it does not offer clear advice about how to control or predict behavior in the future. Patterns that form in a complex adaptive system can be anticipated, but the specific behavior of an individual agent cannot be predicted or controlled. In computer simulation models – which are the primary mode of investigation in CAS – simple rules constrain the actions of multiple agents. Over time, coherent patterns form across the whole.

Figure 1. *Complex Adaptive System* graphically presents the emergent nature of

systemic patterns in a CAS. The random oval shapes at the bottom represent agents of various sizes, dispositions, and characteristics. Over time they interact, and interactions among agents generate emergent patterns, shown as the set of nested ovals at the top of the diagram. When the patterns are established, they constrain the future behavior of agents. You see this in human systems when a culture emerges, then peer pressure causes people to conform to it. The arrow at the left indicates the constraint. Over time, because of the system constraint, the emergent pattern is reinforced and strengthened.

Figure 1: Complex Adaptive System



HUMAN SYSTEMS DYNAMICS (HSD)

CAS and other nonlinear approaches have been used to understand the dynamics of a variety of physical and mathematical systems. Patterned behavior in those contexts is analogous but not identical to emergent behavior in human systems. People are conscious of the behavior of themselves and others, they learn from past experience, and they have hopes and desires for the future that affect their behavior and their expectations of others. Finally, and probably most important, people take intentional action to influence patterns as they emerge. All of these characteristics complicate the complex adaptive nature for humans and the systems they create. For this reason, a literal translation of CAS to human systems is insufficient to help us see and influence patterns that emerge from social interactions.

The emerging field of human systems dynamics (HSD) integrates perspectives of CAS and other nonlinear sciences with traditional social sciences to articulate the complex dynamics that shape self-organizing patterns in social systems at all levels – intrapersonal, interpersonal, within small groups, organizations, and communities (Eoyang 2003). The HSD perspective can inform understanding and action in complex situations, such as the transformation of an organization to integrate services. HSD is based on the definition of social structures as complex

adaptive systems (CAS), but it also incorporates the unique features that humans contribute to the systemic dynamics. HSD shares assumptions with and uses metaphors, tools, techniques, and methods that derive from the new nonlinear sciences, including the study of complex adaptive systems.

In spite of its technical roots in nonlinear dynamics and computer simulation modeling, HSD reflects the deep intuitions and implicit knowledge of wise practitioners who work with people and organizations. It provides an analytical method and shared language that has been used to move individual insights about the complex dynamics of an organization into shared understanding and action.

Why use an HSD approach to evaluate the Cope County reorganization?

The Cope County project involved characteristics that indicated the need for a human systems dynamics (HSD) approach. An HSD approach to evaluation makes sense when:

- participants in the system possess a moderate degree of freedom of action
- systems of interest are defined more by functions than by structures
- change over time is seen as dynamical, rather than dynamic or static
- expectations and outcomes are emergent, rather than predetermined
- system boundaries are open rather than closed
- change is acknowledged at multiple units of analysis (individual, process, organization, client, etc) simultaneously.
-

Let's look at each of these in turn.

MODERATE DEGREES OF FREEDOM

Agents (eg staff, teams, organizations, funders, clients) participating in human systems may have more or less freedom to act in unpredictable ways, and evaluation methods can be selected to reflect the complex and variable levels of constraint that shape emergent patterns and make them more or less effective.

The situation is analogous to the plight of water molecules at various temperatures. Below freezing point, the molecules are tightly constrained. Above boiling point, the molecules are free to move about unimpeded in three-dimensional space. Between melting and boiling points, the molecules have more freedom to move in space, though they usually stay connected to each other. Human systems dynamics demonstrate similar patterns of constraint and resulting systemic behavior.

FUNCTIONS RATHER THAN STRUCTURES

Human systems dynamics (HSD) approaches focus on functions and their interdependencies. Rather than defining specific structures and relationships among those structures (eg formal organization charts), HSD methods include network, ecological, and genomic models in which systemic and functional system features – like patterns – emerge and are subject to fundamental changes in identity over time.

DYNAMICAL CHANGE RATHER THAN DYNAMIC OR STATIC

People who work with systems are often as interested in how change takes place as they are in whether or not change occurs. Physical scientists distinguish among three ways to describe change over time – static, dynamic, and dynamical. Each kind of change requires different systemic evaluation approaches. A static description presents a system at rest. Change, when it comes, is the result of specific interventions that shift the system from one stable point to another. A dynamic description acknowledges change over time, but it assumes a smooth trajectory. Physical examples include the parabolic path of a moving projectile – direction and speed change continually in response to the momentum of the object and the pull of gravity. Dynamical change is influenced continually by variables that are interdependent (rather than independent or dependent). When change in a system is dynamical, the system may shift from rest to rhythmic oscillation to random thrashing. These changes seem to be spontaneous, but they are driven by the internal dynamics of the system itself as the constraining conditions interact with each other to influence the behaviors of agents in the system.

EMERGENT RATHER THAN PREDETERMINED

HSD perspectives assume self-organizing forces within the system emerge from the past into the present to shape future results over time. Results in the future emerge from interactions in the moment. The future is unpredictable because the dynamical interaction of the forces at each point cannot be known in advance.

Clearly, goals have influence over organizational performance and change. This is one of the key ways in which the dynamics of natural, physical complex adaptive systems (CASs) are distinct from human systems dynamics (HSD). Because human beings are conscious and intentional, we can individually envision a future different from the present. Because we have language and social connections, we can collectively align our individual actions to support the goals of the whole. Because of our self-consciousness and connections, we have the means to introduce a shared goal for the future as one of the complex forces that affect our actions in the present. Nevertheless, as participants in complex, highly interdependent and emergent systems, even the most competent individuals cope with unknowable and uncontrollable futures.

OPEN RATHER THAN CLOSED

In his *General Systems Theory*, von Bertalanffy (von Bertalanffy 1980) recognized the differences between systems that are open and closed. Most of his work focused on closed systems, because the questions related to them were more tractable, given the mathematical tools and techniques available to him. Although “closed system” approaches are useful, many systems approaches these days describe systemic behavior in terms of open rather than closed systems.

Complex adaptive systems, and their HSD correlates, are assumed to be open systems. Effective boundaries appear and disappear over time, and are not “given.”

An observer is free to define boundaries for the purposes of description or analysis. In addition to impermanent system boundaries, complex systems are unbounded in two additional ways. They are scaled, so that dynamics at smaller scales (such as individuals) affect patterns that emerge at larger scales (group dynamics). Likewise, larger scales influence patterns at lower levels. Complex relationships among levels of analysis keep complex systems open to unseen influences.

Also, boundaries in human systems are massively entangled. One individual participates in multiple natural systems: work group, family, faith community, and so on. Emergent patterns within each of those boundaries shapes the individual who, in turn, shapes patterns as they emerge in the other contexts.

MULTIPLE LEVELS OF ANALYSIS

Human systems can be conceived to involve massively entangled relationships and multiple levels of interaction. Changes at one level may influence emergent patterns at levels above and below. To capture the dynamic emergence of organizational change, an HSD assessment tracks changes at individual, group, departmental, and organizational levels of scale simultaneously and considers how each of the levels influences the others.

While HSD provides a way to evaluate performance in complex adaptive situations, it is not an appropriate approach when an evaluation design focuses on situations that are tightly constrained or completely unconstrained, on structures rather than functions, on dynamic or static change, predetermined outcomes, closed systems, or single levels of analysis. Many other powerful tools and techniques are available to support evaluations in such situations, so the more “fuzzy” aspects of HSD approaches are not necessary and may even prove to be counterproductive.

In the following sections, we will tell the story of a complex client environment in which HSD approaches were appropriate. We provide conceptual background from CAS and HSD that informed the evaluation design, and describe the method and outcomes of the evaluation.

The Story: Cope County Integrates Human Services

Since the mid-1980s, “integration” has been the “holy grail” for the delivery of public human and social services in the US. The traditional service delivery model was based on bureaucratic isolation of programs from each other. Clients who needed food stamps, child support, job training, and medical benefits had to make multiple connections, endure repetitious registration processes, and pass inconsistent eligibility criteria. Many policy makers believed that integrating these services would not only reduce client frustration but also lower costs and improve outcomes because clients could receive a package of services that responded to their unique needs and challenges.

In spite of a clear belief in and commitment to integrated services and extensive efforts by states and counties, non-profits and business interests during the

1980s and 1990s, few integration programs have been successful. Programs that have succeeded in limited scope or timeframe have proven not to be scalable or sustainable. The experience in Cope County mirrored this national pattern.

In the previous decade Cope County had engaged in a variety of efforts to integrate services, including establishing a pilot of community-delivered services, cross-functional teams, co-locating services, providing information technology to support shared services, and establishing budgets and project teams that crossed program lines to support specific groups of clients. These individual efforts had produced a wide range of success, and none had established long-term or broad-based integration of services. In 2004, the county was ready to make a major organizational transformation to support the integrated delivery of human services.

Cope County is a large, urban county in the Midwestern United States. In January of 2004, six human and social service departments of the county merged into a single Social Services Department (SSD). The purpose of the redesign was to integrate services to improve client outcomes. The newly formed Department included 3,000 employees and took responsibility for a comprehensive list of human service functions, including children, adult, and family services; community health; economic assistance; training and employment assistance; and veterans services. Management and governance structures were redefined, and a consolidated budget was developed. Almost a year later, in the fall of 2004 senior management decided to assess the progress of the redesign effort. In response to concerns from the County Board and Administration, the Senior Management Team of the new Department initiated an assessment to answer the following questions:

- How is the organizational change progressing?
- What recommendations can be made for improving the progress?
- How can we assure that the change supports employees in meeting the needs of clients and communities?

In early conversations with the leadership team, other objectives were defined for the evaluation products and process.

- Develop internal capacity for change evaluation and management.
- Engage an existing cross-functional Redesign Team as partners in the design, implementation, and analysis of the evaluation to form an Assessment Team of internal staff and external evaluators.
- Acknowledge the negative shared discourse about the changes (negative comments were rampant), but focus on behavior and performance issues related to the redesign.

In collaboration with the departmental leaders and a cross-functional Redesign Team, the Assessment Team considered the business goal (integration of services) and the current organizational status. Other evaluation processes were working in parallel with this project to assess client outcomes and service delivery processes. The focus of this project was solely and completely on the organizational redesign.

Given the complex and systemic nature of the environment and the redesign process, we decided that a systemic approach based on HSD principles would be most effective in answering the questions posed for this evaluation.

The next section outlines the tools and frameworks we used to evaluate the reorganization based on HSD principles.

CDE Model for Conditions of Self-Organization

Considering human systems as complex adaptive systems can help us understand in retrospect the patterns that emerged and the developmental path that might have shaped current structures. Such a perspective is not sufficient, however, to shape intentional action that affects human systems dynamics. In order to influence emergent patterns, we need to know and adjust the conditions that determine the speed, path, and outcomes of self-organizing systems. HSD defines three conditions for self-organizing in human systems: C, D, and E (Eoyang 2001).

First, a container (C), bounds the system and determines the sub-set of agents that will interact to form collective patterns of systemic interest. In the Cope County case, we can consider a variety of factors that function as containers. Each management level is a boundary of sorts and generates patterns of systemic behavior over time. Individuals at all levels of the organization hold their own histories and identities, making each of them a relatively contained, self-organizing structure. Each change initiative executed during the preceding year functioned as a relatively independent, emergent system. And ultimately, the newly formed SSD functioned as a powerful organizational container, in which the desired pattern was greater integration of services for clients.

Any human system includes an unknowable number of containers. In Cope County, for example, other containers might include professional groups, cultural or racial identity groups, groups with shared history, those with seniority, pay grades or scales, and so on. Luckily, dealing with all possible containers is neither possible nor necessary. An HSD approach focuses on the most relevant emergent patterns and the containers within which those patterns emerge.

Second, differences (D) within the container establish the tendency toward motion and define the features of the pattern that emerges across the system. Difference is the engine that drives self-organizing behavior. Without difference within a container, nothing will happen – entropy rules. Distinctions among schedules and eligibility requirements drove the need to integrate services, and the variety of internal policies and procedures set the conditions for the redesign of Cope County. At the same time, differences serve the function of articulating the patterns that result from self-organizing processes. If the organizational transformation is successful, Cope County will not be a homogeneous whole. Rather, it will incorporate systemic patterns of differentiation that fit the environment, including the needs of various client groups, demographic trends, fiscal constraints, and political and public expectations for services. Each of the Cope County containers (management levels, individuals, work groups, change

initiatives, and the Department as a whole) includes its own set of significant differences. For example, within the management level container, one would expect funding to be a significant difference, but this would be less significant in the self-organizing perceptions of an individual worker.

The third, and last condition for self-organizing describes the interactions or exchanges (E) among the agents. Exchanges provide the interactive options that allow system-wide patterns to emerge. Traditional systems dynamics approaches deal primarily with patterns, as they model flows and feedback loops². Forms of exchange, like differences, vary from one container to another. For example, we considered decision making as the most relevant exchange between and among management levels for Cope County. At the individual level, however, we collected statements of personal perception and values as exchanges that articulated emerging patterns of personal meaning.

HSD deals with this triad of determinants as the conditions for self-organizing in human systems – the CDE Model. This model is useful both in seeing patterns and influencing them as they emerge because the three conditions are interdependent. A change in one condition results in a change in the other two conditions. As the conditions change, future patterns are transformed. For example, increasing the size of an organizational container (as Cope County did by creating the mega-department) weakens the exchanges and increases the number of differences in the pattern. In the larger container of the new department, old relationships are disturbed and new connections are formed. Because the new exchanges are more numerous and cross a wider range of concerns and interests, individual and systemic behavior are less constrained over time, and emergent patterns are slower to form and less coherent.

Though the CDE Model can provide insight into relationships, emergent patterns, and options for action, it cannot be used to predict or control systemic outcomes. The reason is simple. The system is much more complex than any specific CDE description can capture. A systemic description focuses on a small number of containers, differences, and exchanges that appear most relevant to the patterns of interest. In reality, however, conditions considered irrelevant to a particular description can (and frequently do) disrupt anticipated systemic behavior. Toward the end of the Cope County project, for example, the County Board and County Administrator became interested in the assessment project. They were disappointed that we had not included community input in the evaluation process, though the community was not a container that we considered in our project specifications or ensuing design. This shift in interest affected how our findings were implemented and subsequent projects in ways we could not have predicted or controlled.

2 See the contributions from Ken Meter, Richard Hummelbrunner, Jay Forrest, and Dan Burke.

HSD Assessment of Organizational Change

The Cope County redesign process was complex and emergent, matching the assumptions and conceptual criteria of CAS and HSD. Though a variety of design approaches were available to us, from structured program evaluation through traditional systems dynamics, we chose to base our assessment design on the CDE Model and to explore the self-organizing conditions and patterns in the multiple complex systems embedded in SSD. This section describes the design, findings, recommendations and actions related to the Cope County redesign assessment.

ASSESSMENT DESIGN

Because the C, D, and E are causally connected to each other, assessment questions about one of them can reveal information about the state of others. Our assessment of the organizational change of Cope County considered five primary containers: Management levels, individual employees, service delivery processes, change initiatives, and the Department as a whole. Within each of these containers, we explored either the exchanges or the differences to gather information about the other conditions and to see the emergent patterns in each level clearly. Seeing the patterns in terms of the CDE allowed us to identify ways in which the patterns of organizational change could be more effective and efficient.

The design included five data collection and analysis activities to answer five distinct questions – one about either the differences or exchanges that formed patterns within each of the key containers involved in the organizational change. The sequencing of the activities is somewhat arbitrary. They could have been completed in any sequence. The activities had no natural sequence because the patterns of interest were emerging continually, and it would be impossible to predict how any one of the activities might influence subsequent activities or the patterns that might be revealed.

Vertical Alignment: How well are management levels (C) aligned in the factors they consider in decision-making (E)?

This activity focused on exchanges within the container of levels of management. The purpose of this activity was to provide insights into the similarities and differences among managers at each level in the organization and between management levels. These connections – between and among levels of power within an organization – provide the capacity to support individuals and processes as an organization undergoes change. Seventy-nine persons from four management levels participated in five focus groups to consider the factors that affected the decisions they made.

Networks of Meaning: What are individual employees (C) saying about their experiences of the shift to one department (D)?

This activity articulated the differences among individuals and revealed insights into their methods and modes of exchange. In this activity, each employee of SSD was invited to voice insights and concerns by submitting open-ended response to

two questions: “What’s working in the shift to one department?” and, “What’s not working in the shift to one department?” These questions focused attention on two critical differences – working/not and before/after. The essays invited any staff member in the organization to share their own language about what was working for them, the organization, and their clients, as well as what was not working. Of 3,000 employees, 736 submitted responses; most of those were received by anonymous email. Patterns of response were identified by manual content analysis. In addition, essays were analyzed with CRAWDAD, an automated tool that uses centering resonance analysis to define networks of meaning in text.³

Horizontal Alignment: How do the different service delivery processes (C) of the Department connect with each other to provide service (E)?

This activity explored exchanges between and among work groups as containers to identify the different levels of integration and how integration activities had changed over the previous year. This component elicited information about how different work groups across the Department are interacting with each other on a daily basis. Such interactions are the core of integration of services, and the data collected provided a portrait of integration that can be used to identify options for action and to evaluate integration as it progresses. Ten SSD staff were trained in the data-collection protocol and worked as paired facilitators. One hundred, sixty-two staff members participated in a total of ten focus groups. Data was collected in graphic form as each participant indicated input and output connections for their own work processes, identified interactions that had changed in 2004, and whether the effect of those changes had been positive, negative, or as yet undetermined. The graphic data was converted to a spreadsheet that indicated all of the interactions, changes, and effects of the changes across the Department, the county, and outside of the county government.

Common Language of Change: How do change initiatives (C) define their work to implement specific changes (D)?

This component focused on the differences between and among change projects across SSD and stimulated new exchanges between and among individuals and projects across the Department. This activity brought together change agents from across SSD to share their insights, provide an understanding of the types and amounts of change activity currently underway, learn a shared model for the process of emergent change, and celebrate their accomplishments. Rather than arbitrarily defining “change leaders” or centrally selecting persons to be included in the activity, we asked managers and supervisors to recommend persons who had been involved in or led a change-related project over the previous year. Everyone who was recommended was invited to the event. One hundred individuals attended the one-day session. Data collected included a list of over 150 change-related

3 CRAWDAD was developed by Drs. Steven Corman and Kevin Dooley of Arizona State University. Information about the tool is available at <http://www.crawdadtch.com/software.htm>.

projects, challenges faced by change agents, and opportunities to improve change processes in the future.

Internal Documents: What do system-wide communication documents (E) reveal about the process and depth of the change across the Department (C)?

This activity explored the exchanges across the Department over the previous year to identify key differences and recommend future communications strategies. This activity examined the ways in which SSD leadership and staff had communicated with each other in writing about the redesign efforts, the integration of services, and the improvement of outcomes for clients. One hundred, seven clusters of related documents were reviewed, including official management communications, meeting minutes, newsletter articles, and change initiative reports. The source and audience of each communiqué, in addition to its timing and core messages, were analyzed.

ASSESSMENT FINDINGS

Each component of the assessment generated a rich source of qualitative data about the patterns emerging from the redesign efforts and about the conditions that were or were not facilitating coherent change across the system. At the conclusion of each activity the evaluators completed a preliminary analysis. Summaries of data were then presented to the Redesign Team, who further refined the analyses and provided interpretations from their perspectives of the context and history of the organization. A detailed report of each component was prepared to reflect the data and analysis. The detailed reports were presented to the Executive Committee at the conclusion of each activity, and interim actions were defined and recommended for immediate implementation based on the findings from each component. Finally, the detailed reports were posted electronically so that all staff had access to the findings from each of the assessment activities. At the end of the project a summary report was produced and distributed.

Through these means of documentation and broadcast reporting, the project not only assessed the conditions (CDE) for self-organizing, it also helped shift the conditions toward different patterns by establishing new containers, differences, and exchanges across the Department.

WHAT PATTERNS EMERGED?

Similar patterns appeared to some extent in each one of the assessment activities – management levels, individual employees, service delivery processes, change initiatives, and the Department as a whole. Given the similarity of the patterns, one might ask whether assessment of all five were necessary. Though the systemic patterns could be discerned after the fact in data collected from patterns within each container, we believe that the five-part design is preferable for a variety of reasons. First, multiple activities allowed the team to triangulate the patterns as they emerged, so a shorter list of more significant patterns could be articulated clearly. Second, each activity involved more and different staff members as

integral parts of the assessment effort. This increased sense of engagement was a key outcome of the assessment process. Third, not all of the identified patterns appeared equally strongly in all of the activities. Depending on the subset of activities included, one of the major patterns might not have appeared to be significant. Finally, the redesign effort was viewed from a variety of quite distinct perspectives. Data collected and analyzed from any one of these would have been suspect. This diverse and broad-based design helped establish credibility for the findings and commitment to the recommendations.

Across the multiple data sources and analysis processes, the following patterns emerged as critical to the ongoing success of the organizational change effort:

- Many exciting changes in service process, delivery, and outcomes are already being realized from the redesign.
- Staff members understand and support the vision of integrated service delivery and improved outcomes for clients, but they are anxious and confused because they do not have a clear picture of the path that will lead the Department into this different future.
- Individuals across the organization are feeling the natural discomfort related to large-scale organizational change. Structural and staffing changes have left individuals feeling disconnected from each other and from the organization.
- This new organizational form – beyond silos – presents a new landscape of opportunities and accountabilities. Managers and staff at all levels are unclear about how to develop skills they need to be successful in this new way of working.
- An agile organization requires efficient decision-making processes that use knowledge from all levels of the organization. Today in SSD, decision-making is slow, concentrated at the top, and not transparent. No process exists to resolve conflicts so we can move ahead together. As a result, the tremendous adaptive potential of SSD staff has not yet been realized.

Other kinds of evaluation methods might have identified one or more of these patterns using approaches that were not explicitly based on HSD principles. Experienced organizational practitioners see patterns of human systems dynamics intuitively and take or recommend action to respond to what they see. The HSD approach is distinct from standard consulting practice and insight in at least four ways. First, the CDE assessment design made the emergent patterns manifest, so that everyone involved in the project was able to see and name them. Second, participants saw underlying dynamics that could inform their work in a wide variety of contexts. Third, the patterns were captured at multiple levels of analysis (management levels, individuals, service delivery processes, change initiatives, and Department-wide). Fourth, HSD describes dynamical relationships that lead to recommendations for action to shift self-organizing patterns as they emerge.

ASSESSMENT RECOMMENDATIONS AND ACTION

Not only does an understanding of complex adaptive systems in a human context reveal systemic patterns of interaction, it also affects options for action to improve those patterns. One of the challenges of a complex system is the wide array of possible interventions. Individuals or groups can influence change. Managers and staff can make a difference. Each work group can take (sometimes opposing) action to move toward new goals and objectives. The multiplicity of possible actions can be overwhelming and confusing, and when individuals are overwhelmed and confused it is difficult for them to take aligned and coherent action.

CAS, and its application to people in HSD, provides a simple way to influence coherent action across complex and diverse human systems. The idea derives from computer simulation models in which a large number of agents generate coherent system-wide patterns when all agents follow the same short list of simple rules. The classic example is the cellular automaton called BOIDS.⁴ Simulated birds move around on a computer screen according to three simple rules.

1. Fly toward the center.
2. Match the speed of the flock.
3. Don't bump into others.

When all of the system's agents follow these rules, the emergent systemic pattern of behavior looks like a flock of birds moving in coherent and recognizable patterns. Applied to human systems, this concept of a short list of simple rules (Shorts and Simples) can help a very diverse group respond to unique concerns and issues and still work together to generate coherent system-wide behaviors.

We used the concept of simple rules to transform the findings of observed patterns into recommendations for Cope County. Table 1: Framework for Action is the format in which we presented our summary findings and recommendations to Cope County personnel. It shows how the findings of observed patterns suggested a simple rule, and how each simple rule suggested specific actions for the Executive Committee.

In addition to system-wide recommendations for the Executive Committee, every individual and subgroup within the Department was encouraged to consider the same short list of simple rules and identify actions that he or she could take to work in consonance with others and with the Department as a whole. This recommendation reinforced the notion of the complex adaptive system, as each employee and work team was encouraged to be conscious of its own participation in an emergent, Department-wide pattern.

The first assessment activity in the project began the first week of January 2005, and the project concluded in mid-April 2005. Many changes have been made to implement the recommendations from the assessment and to put other lessons

4 Various versions of BOIDS are available on the Internet. This one is particularly interesting and easy to manipulate: <http://www.red3d.com/cwr/boids/>.

Table 1: Framework for Action		SSD April 2005
Observations	Simple Rules	Recommendations for Executive Committee
<p>Communicate Many exciting changes in service process, delivery, and outcomes are already being realized from the redesign.</p>	<p>1. Build success for yourself and others.</p>	<p>a. Talk about strengths and accomplishments to encourage effective action toward integrated services and improved outcomes.</p> <p>b. Provide clear boundaries for staff decision-making and action, then accept solutions they develop.</p> <p>c. Take action and communicate regarding recommendations from existing groups (eg Governance Grid, Data Sharing, Front Door, Performance Management, Recognition).</p> <p>d. Be gentle with yourself and others. Remember we all have feelings.</p>
<p>Plan for the whole Staff members understand and support the vision of integrated service delivery and improved outcomes for clients, but they are anxious and confused because they do not have a clear picture of the path that will lead the Department into this different future.</p>	<p>2. Develop people and processes that improve outcomes.</p>	<p>a. Adopt and implement a comprehensive, strategic communication plan to build shared story about the strategic direction.</p> <p>b. Provide a road map for the phases and stages of the redesign transformation process.</p> <p>c. Improve what you do and how you do it by establishing and documenting effective processes and infrastructure.</p>
<p>Overcome isolation and fear Individuals across the organization are feeling the natural discomfort related to large-scale organizational change. Structural and staffing changes have left individuals feeling disconnected from each other and from the organization.</p>	<p>3. Stay connected.</p>	<p>a. Hold all who manage others accountable for meeting with their direct reports regularly to share strategic direction, answer questions, and to listen and respond to concerns.</p> <p>b. Publish a current directory of contacts within SSD and an org chart through supervisor level. Keep it up to date.</p> <p>c. Be clear about and accountable for performance expectations.</p> <p>d. Share what you know, including your questions. Don't wait for perfection. News will never be perfect until it is shared.</p>
<p>Build competencies to support integration This new organizational form – beyond silos – presents a new landscape of opportunities and accountabilities. Managers and staff at all levels are unclear about how to develop skills they need to be successful in this new way of working.</p>	<p>4. Learn your way into a shared future.</p>	<p>a. Provide personal coaching, training, and support to help directors, managers, supervisors, and staff work most effectively in the SSD of tomorrow.</p> <p>b. Hold all who manage others accountable to establish the conditions for creative engagement at every level across the Department.</p> <p>c. Define, implement, and enforce Service Level Agreements between Internal Support groups and their clients.</p> <p>d. Implement the Balanced Score Card and other dashboard measures across the whole department.</p> <p>e. Regularly and frequently review progress of the redesign effort and take steps to adapt.</p>

from the assessment to work in various parts of SSD. Most significantly:

- An employee satisfaction survey is structured to measure performance on each of the five simple rules.
- The Executive Committee has implemented boundaries for staff decision-making and clearer definitions of roles and responsibilities across the Department and between management levels.
- A strategic communication plan has been developed and execution has begun.
- A road map for change is being defined for distribution.
- A directory of contacts for various service areas is available on the intranet.
- Service Level Agreements and a process for acceptance and monitoring have been designed and implemented.
- Balanced Score Card and other feedback procedures are being implemented across the Department.
- Many employees at multiple levels have considered the Shorts and Simple and identified ways in which these simple rules might shape their own action.

Plans are in place to repeat some portions of the HSD assessment design on an annual basis to continue to track the emergence of system-wide patterns.

Conclusion

Integration of public social and human services is a worthy cause and promises to improve client outcomes while controlling costs and improving worker satisfaction. The path toward service integration, however, requires individuals and organizations to enter into a realm in which boundaries are unclear, criteria for success are transformed, and relationships are as critical to success as they are strained by changing expectations.

Under such conditions, human systems enter into a regime of highly emergent and systemic pattern formation. If the organization had been in a more stable, predictable state, or if our focus had been on specific, measurable outcomes, HSD would not have been our choice for an evaluation approach. When an organization is in an active state of complex adaptation, however, traditional evaluation techniques may be insufficient to capture the critical factors that shape the current and future performance of the system. We recognized the patterns of CAS at Cope County, so we chose an HSD approach to answer the evaluation questions about this systemic, organizational change.

In today's fast-paced and highly interdependent organizations, innovative evaluation strategies are sometimes required. Cope County and its HSD-based assessment project demonstrate how knowledge of complexity and human systems dynamics and of their implications for individual and organizational performance can offer options for understanding complex evaluation situations.

References

- Bertalanffy, L von. 1980. *General systems theory: Essays on its foundation and development*. New York: George Braziller Publishers.
- Cowan, G, Pines, D, and Meltzer, D. 1994. *Complexity: Metaphors, models, and reality*. Reading, MA: Addison-Wesley Publishing Company.
- Dooley, K. 1996. A complex adaptive systems model of organizational change. *Nonlinear Dynamics, Psychology, and Life Sciences*, 1(1): 69-97.
- Eoyang, G. 2001. *CDE model: Conditions for self-organizing in human systems*. Unpublished doctoral dissertation. Cincinnati, Ohio: The Union Institute and University.
- Eoyang, G (ed). 2003. *Voices from the field: An introduction to human systems dynamics*. Circle Pines, Minnesota: Human Systems Dynamics Institute Press.

USING SYSTEMS CONCEPTS IN EVALUATION

Evaluating Farm and Food Systems in the US¹

Kenneth A Meter

Part way through Ken's chapter you may wonder what an elegant critique of mid-20th century agricultural economics has to do with systems and evaluation. Quite a lot as it turns out. For a start it highlights the importance of history in the systems field – something often downplayed in many systems-based approaches. Vital clues to our mental models are to be found in the past. Ken then picks up on Gerald Midgley's idea that greater insights can be gained from a situation by using multiple methodologies. In this case, he draws on methods from four methodologies already described in this volume; system dynamics, soft systems, complex adaptive systems and critical systems. The use of critical systems may not be immediately obvious, but consider this – isn't using "wise elders" an application Critical Systems Heuristic's use of critical expertise to question the dominant consensus ?

Summary

Evaluation of food and agricultural system activity in the US is complicated by systemic economic relationships that extract considerable wealth from rural communities. These external pressures severely limit the options available to community-based food systems initiatives, and may confuse evaluation efforts. Analysis of regional farm and food economies, informed by the insights of "wise practitioners," illuminates the nature of these extractive relationships, setting the stage for more precise systems evaluation. In this paper, economic data will be applied to three systemic evaluation methods, drawn from Systems Dynamics, Soft Systems Methodology, and Complex Adaptive Systems. How well does each of these methods account for available data? How might each be used? How might evaluators use these tools to engage in systems analysis, or to devise or assess progress toward specific theories of change? How can insights from wise practitioners be tested and incorporated?

Evaluating farm and food systems in the US

LOCAL FOOD AND FARM SYSTEMS

A dynamic, diverse movement in the US now attempts to build community-based food systems (CBFS) in thousands of urban and rural locations (Meter 2003). As one example, a cluster of 50 growers and producers have formed the Southeast Minnesota Food Network. Here, farmstead butter and cheese makers, large-scale orchards, food distributors, coop retailers, and specialty meat producers

¹ This paper draws upon insights graciously provided in reviews of early drafts by Bob Williams, Bill Harris, Lee Mizell, Glenda Eoyang, Tom Berkas, David Scheie, and JoAnne Berkenkamp. Additional insights were offered by Martin Reynolds, Richard Bawden, Gerald Midgley, and other authors in this volume, during a meeting in October, 2005.

collaborate with small community-supported agriculture (CSA) produce farms (in which consumers buy food “shares” in advance), linking businesses into a coordinated effort.²

That such activity takes place within an advanced farm economy with such high apparent productivity is remarkable. That its proponents view themselves as creating new food systems requires evaluators to apply systemic evaluation techniques.

Local activities are complex in themselves, yet they are deeply impacted by complex, global relationships (such as global commodity markets and capital flows) that are difficult to understand. No matter how one might wish to simply draw a boundary around local action and limit one’s work to that setting, this often proves impossible.

Similarly, selecting appropriate evaluation tools can be a daunting task. Evaluators may play a variety of roles in any given assessment, including framing, revising, or measuring progress toward a theory of change. They may be called upon to interview participants, or to summarize survey responses or other quantitative data. Evaluators may be the strongest voice upholding a long-term or systemic vision. Systems methods described here may all have utility for any of these tasks.

This evaluator’s professional experience has shown that systemic evaluation efforts are often hampered by (a) the difficulties of modeling systemic activity concisely; (b) a lack of understanding of economic constraints; and (c) overlooking the insights of important stakeholders. Thus, this paper begins with an overview of economic lessons that have emerged while evaluating food-systems activity. Then it will show how diverse evaluation methods may be applied, especially in modeling, by incorporating economic analysis that offers simplifying insights, views from multiple perspectives, and testimony from wise practitioners.

For this paper, CBFS are defined as *systems of exchange that strive to bring food producers and food consumers into affinity with each other*, for the purposes of fostering health, promoting nutrition, building stronger community ties, keeping farm families on the land, and building wealth broadly among community members (Meter 2003 p8).

This contrasts with the prevailing US agricultural economy, which focuses on production of *commodities* that are more typically raw materials for further processing than actual foods to be eaten directly by humans. Less than half of one percent of all U.S. farm commodities is sold directly to consumers.³

Both community-based and commodity-based food systems have interacted on the North American continent since Europeans first settled here. Shifting economic and policy winds have altered their relative strengths. Comments by several “wise practitioners” led to key indicators that measure this everchanging balance.

2 Further information about the network can be found at <http://www.localfoodnetwork.org/>

3 US farmers produced \$201 billion of commodities in 2002, of which \$812 million (0.4%) was sold directly to consumers (USDA/NASS Agricultural Census 2002).

WHAT DOES A HEALTHY FARM ECONOMY LOOK LIKE?

In the late 1970s, working as a journalist to cover the impending depression in the farm economy, I asked a group of Minnesota farm neighbors how they could tell when the farm economy was healthy. Without using the term, and long before I worked as an evaluator, I had asked the farmers to suggest an indicator.

The men replied without hesitation, thinking back to the days, twenty-five years earlier, when they had started farms in this community. They told me that when their farm economy was strong, their rural community had its own supply of credit, sufficient to cover the costs of farm production (Meter 1990).

In those days, any farmer worth his salt could – and was expected to – earn the money to make a down payment on land by simply starting to farm. Food these farmers raised was largely consumed locally, through commercial channels that were relatively farm-friendly. Farmers received a greater share of the retail price of food. One man in this circle raised eggs for a year, bringing in enough profit to make a down payment on land the next. Another invested savings he held from a previous farm. Others might ask a parent, or other relative. Only as a last resort would a farmer visit the local banker for a loan. “Back then, it was like a sin to borrow money,” the farmers added with one voice. To them, paying interest – even to a local bank – meant taking money out of the farm community (Meter 1983 p3).

A quick look at farm credit data confirmed their tale. Aggregate farm debt was \$6 billion in 1950, during the time they were describing. By 1985, farm debt had soared to \$222 billion (Meter 1990 p8–9),⁴ – and it had become clear farmers would not be able to pay it back. Tracking the unsustainable debt loads their neighbors carried, these farmers told me there would be a farm crisis soon. Their prediction was entirely correct.

TESTING THESE SOURCES

I trusted the stories told by this group of wise practitioners, not only because federal data bases confirmed their stories, but also because their stories passed severe internal tests. Accustomed to meeting together in the context of a local environmental action group, this cluster of farmers had raised children, shared farm chores, and weathered crises together. Any story was subject to close scrutiny from others in the group. Incorrect notions met a quick challenge. Running diverse farm operations and holding differing skills and needs, these farmers brought varied views to each conversation.⁵

Moreover, the farmers’ testimony was persuasive because these men were both immersed in information about the impending crisis, and detached from it. None of this group had applied for the large loans that typified farm lending at

4 Data drawn from USDA Economic Research Service, Farm Income Statement and Balance Sheet. Recent data from this series is available at <http://ers.usda.gov/Data/FarmIncome/finfidmu.htm>. Current value of this debt (in 2005 dollars) is \$403 billion.

5 Of course, it is also possible for consensus to obliterate the truth, or to marginalize important but unpopular views. In this case, the consensus seemed to this viewer to accompany a sense of openness to new evidence, rather than a closed interpretation, but of course this is a subjective determination.

the time – and which, the farmers correctly surmised, could not be repaid. Most of the farmers in the group I spoke with had, moreover, served on the local county committee for a federal loan program. Scrutinizing their neighbors' farm business plans, they knew the fundamental economics intimately. They had seen their neighbors succumb to lenders' pressure to take on more debt than they wanted, and they knew how their neighbors felt about that.

As it turned out, the farmers' comments led me to decades of research and writing. Their practical experience attuned them to indicators that economists from USDA and other federal agencies had overlooked.⁶ Moreover, by following their intuitions and by refining their analysis through informal discussion,⁷ they had surpassed the ability of federal agencies to understand the impending crisis. They also had more freedom than official experts to report their conclusions, facing few of the political pressures that are routinely placed on academic researchers and agency staff.

Now, after extensive follow-up research over 25 years, it is clear to me that the indicator they chose – the strength of responsive local credit sources – is indeed a profound measure of the health of farm communities. While actually compiling such data is extremely impractical due to privacy concerns, widely reported surrogate data provide compelling evidence that confirms the farmers' views.

CONSULTING THE DATA

What USDA's Economic Research Service does report is the amount of farm debt held by "individuals and other" lenders. This is available for each year since 1910. While not identical to "local" credit, it overlaps powerfully.⁸ Individual lenders were the primary source of farm debt from 1910 until 1972, with the exception of the New Deal years, as can be seen in chart 1. In the early 20th Century, when three of every four dollars of farm debt was held by individuals, most loans would have been held by relatives or neighbors, simply because rural economies were more localized.

Now, nearly a century later, with greater capital mobility and widely dispersed families, it is harder to equate the two. Yet it is still true that all individual lenders are "local" in a meaningful way – each loans money to farmers for reasons that are not strictly commercial. Each is part of the farmer's community, rather than strictly an entry on the balance sheet. Individual lenders offer credit which is responsive to farmers in a different way – perhaps more strict or perhaps more lenient – than for commercial or public lenders.

6 I discussed these findings with a retired senior ERS economist at a national conference in 1995. His comment: "Wow. We never even thought of anything like that."

7 See also Flood 1999 p68-69.

8 It must also be stressed that the term "community-based" (or perhaps better, "responsive") credit may be more useful to apply here than "local." Certainly a sharecropper who was forced to borrow from his landlord in order to raise a crop would not herald "local" credit as an ultimate goal, since this would perpetuate his dependency. For these white farm owners of Minnesota, without direct experience of such exploitive practices, and who could assume some overlap among "community," "responsive," and "local," this distinction had not yet been addressed.

This is not the place to demonstrate the above story in greater detail, but one chart showing farm credit sources is shown below. For our purposes it is enough to state a few select facts. Two eras in US farm history are generally recognized as the peak times for agriculture: the “golden era” of 1910–1914, and the post-WWII expansion which lasted until the mid-1950s. During each era, individual lending peaked. As mentioned, 75% of all farm debt was held by individuals during the earlier golden era. Later, in the postwar period, fewer than half of farm loans were extended by individuals – yet this represented a recovery in individual lending, after 1933, when New Deal policies restored to rural communities their capacity to lend. In both eras, foreign markets were strong, and urban populations were expanding. Credit was not sufficient by itself to cause this prosperity, yet prosperity was based upon the responsiveness of local credit sources to farmers who wished to reach expanding markets.

However, during two severe agricultural crises, individual lending plummeted. Lack of credit was a significant cause of the 1920s farm depression, a global crisis which, many experts have concluded, was the precipitating factor for the Great Depression.⁹ Few Americans realize this was a global agricultural depression (Rothermund 1996). Ultimately, under the New Deal, the federal loans that were extended to farmers worked to restore local savings and thus, community capacity to extend credit.

Farm Debt by Source, 1910 - 2002

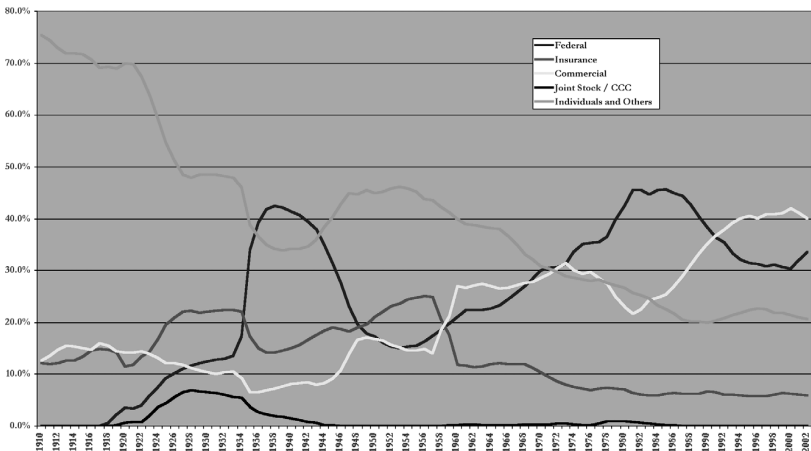


Chart 1

CBFS are mainstream

CBFS decline

Old CBFS fade

Farmers choose local or export

exports dominate

New CBFS emerge

9 See Galbraith (1954); Timoshenko (1933); Friedman and Schwartz (1963); Temin (1976); Perkins, (1969); van der Wee, (1972); Martinus; and Latham (1981).

Similarly, the 1985 farm debt crisis – which the Minnesota farmers I interviewed were predicting – was instigated by the “grain for oil” trade in 1973, in which farmers were asked by the federal government to export large quantities of grain to compensate for rising oil costs during the OPEC oil embargo. Farmers complied, and federal lenders pressured farmers to take on larger and larger loans to expand production. This created a short-term windfall for many farmers, but also encouraged many to take on debts they could not repay. Moreover, the new technology these farmers adopted, enabled by federal loans, was too expensive for local lenders to support.

A few years after the peak of the crisis, individual farm credit had fallen to 20%, its lowest level of the century. Individuals had become the third most important source of farm debt. Unfortunately in this more recent crisis, federal intervention *weakened* rural credit sources, in favor of commercial and federal lenders. Thus, these loans worked, over the long haul, to undermine the very foundation of the rural economy. Farmers began to make more and more of their interest payments to distant lenders who were both unresponsive, and unlikely to reinvest in farm communities.

THE EXTRACTIVE ECONOMY

In fact, from 1913 to 2005, US farmers paid \$595 billion more in interest payments than they received from federal farm subsidies.¹⁰ This means that farmers have subsidized the mainstream economy. Moreover, as credit markets became increasingly global, farm interest payments increasingly failed to recycle back to farm communities. Potential investment capital was drawn away from the rural communities in which farmers produce commodities. The potential for rural regions to build wealth of their own has been weakened.

Thus, farmers operate within an economic context that is increasingly very efficient in extracting wealth from rural communities, and very inefficient in building wealth in those locales where primary commodities are produced.

This data suggests that CBFS stand as a self-organized counterpoint to the prevailing extractive economic structures. From the point of view of citizens in communities, the global economy exhibits considerable disorder. Yet this “disorder” is actually the outcome of a lack of power amidst highly structured global relationships, dominated by international firms that hold considerable influence over global markets, concentrated investment capital, high levels of technology, with significant barriers to entry. These conspire to create immense power for those who command these systems. At the same time, they enforce tremendous disinvestment and deep powerlessness for those who do not.

¹⁰ Calculated by the author using constant 2005 dollars from USDA Economic Research Service, Farm Income Statement and Balance Sheet data. Recent data from this series is available at <http://ers.usda.gov/Data/FarmIncome/finfidmu.htm>.

So what?

Evaluation and the use of systems methods

Knowing the farm economy is extractive – that more wealth is removed from, rather than retained inside, producer communities – is critical to effective evaluation of community-based food system activity. Many of the difficulties encountered by community foods initiatives derive from this extractive character. Next, we turn to how evaluators may make use of this understanding.

THREE MODELING METHODS

Each of the three modeling tools will now be outlined briefly. Each may be applied in building logic models for CBFS initiatives. Practical considerations for using each will be assessed.

CAUSAL-LOOP DIAGRAMS (CLD)

The most visible proponent of causal-loop techniques is Peter Senge, who popularized this System Dynamics approach in his best-seller, *The Fifth Discipline* (Senge 1990). Senge focuses on building a consensus among diverse stakeholders so they can agree how best to implement a strategic plan (Flood 1999). Senge further identifies “system archetypes,” representing classic processes in which delays or feedback make the outcomes of a given action difficult to foresee. If people recognize these archetypes in their local food systems work, it may bring greater unity to their understanding of the systems issues they face.

SOFT SYSTEMS METHODOLOGY (SSM)

“Soft” systems methods were developed by Peter Checkland and others to address issues that arise without clear, tangible definition or boundaries. Often in such cases, straightforward numeric analysis is difficult. Multiple perspectives are likely to be valid. SSM explicitly distinguishes “real world” phenomena from the systems model itself. Collaborators work together to model a systems condition they face, and then compare that model (or often multiple models) with actual events.¹¹

COMPLEX ADAPTIVE SYSTEMS (CAS)

A number of Complex Adaptive Systems approaches focus on the complex and changing nature of systems, recognizing that as people within a system take action, the system itself changes. One qualitative version of CAS suggests that those who launch a systems change initiative perform an analysis of CDE: Container, Differences that make a Difference, and Exchanges.¹² CDE attempts to

11 Much of the information covering SSM is drawn from Williams 2002 and 2005. See also Boon Hou Tay's and Kate Attenborough's chapters.

12 Material covering CAS is drawn from Eoyang & Berkas (1998) as well as Eoyang (2004). Also Williams (2002) and Williams (2005); Lichtenstein (2000). Note that the Evolutionary Agroecology and Biocomplexity Initiative at Iowa State University defines agriculture as a complex adaptive system. See also EABC Statement of Terms. Australian researchers identified emergence as a key property of food systems, eg Crawford, Nettle, Armstrong and Paine (2002). Although not specifically analyzed by Sørensen and Vidal (2002), CAS approaches appear to address their concern that SSM and other soft approaches require expert facilitation. See also Glenda Eoyang's chapter.

assist those working within complex systems to define that changing context, and to assess the efficacy of their efforts given the changes that occur.

Each approach has its particular strengths and weaknesses. These will be characterized according to the following qualities (see summary chart on page xx):

- Easily understandable
- Expresses feedback and other systemic qualities
- Heuristic value (leads to future learning)
- Expresses separation between “reality” and “model”
- Lends itself well to lay use
- Builds agreement among diverse stakeholders
- Designed to build consensus among participants
- Expresses change over time
- Embraces multiple perspectives
- Expresses power dynamics at work within the initiative itself
- Designed for use in a highly bounded, stable or specific organizational context
- Expresses stocks (accumulation) and flows (movement) of resources
- Lends naturally to measurement of key dynamics

Now, we turn our attention to each of these respective modeling methods.

CAUSAL-LOOP DIAGRAMS AS A TOOL

Many groups use causal-loop diagrams as a tool for strategic planning. These diagrams are a significant advance over more linear models often used in theories of change, since they account for the ways in which systems will “push back” against (resist), or reinforce, efforts to change the system.

Focused on work within organizations, Senge argues in *The Fifth Discipline* for making explicit the feedback loops that tend to either amplify or offset any activity that may be initiated, distilling these understandings into archetypal diagrams.

Many practitioners caution that it is most fruitful to use causal-loop diagrams to model a *change initiative*, rather than the *system itself* (Eoyang 2005), simply because this creates a more concise image of the proposed change and potential resistance to it. Using causal-loop diagrams to model entire systems can lead to large, unwieldy charts.

One of Senge’s archetypes, “Limits to Growth,” could be applied to CBFS movement, showing how efforts to increase the scope of CBFS activity may be frustrated by systemic pressures (Senge 1990 p379). The diagram below is one such way of modeling this scenario.

The downward curving arrow on the top right represents potential donations or investment by foundations or businesspeople. Investors may seek rapid financial return, or may invest for more patient, long-term benefit.

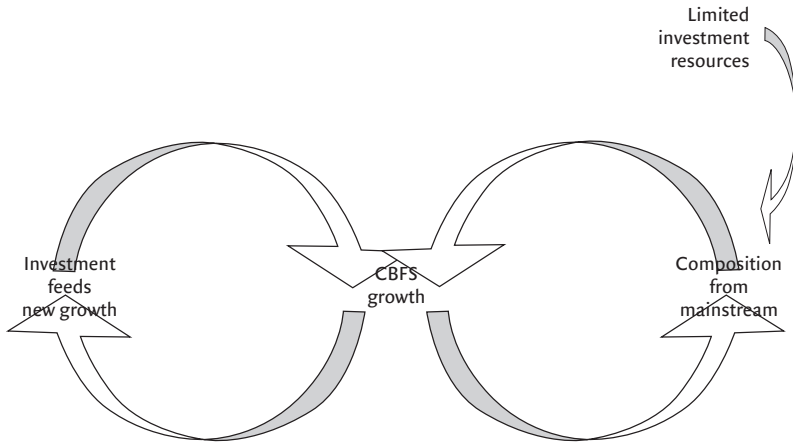


Figure 1: Causal-loop diagram of limits to growth archetype

Note: in this causal-loop diagram, “R” signifies feedback that reinforces the action being initiated, while “B” signifies feedback that opposes or “balances” this action.

Investment, of course, could also be channeled directly to the CBFS itself, which could be represented by an arrow pointing to “CBFS growth” at the center of the diagram.

Such a diagram may serve to unify the vision of participants in a CBFS initiative. Used in an iterative process, in which succeeding diagrams refine earlier ones, even a simplistic CLD may have heuristic value, by identifying critical stages where unintended impacts may occur. This strengthens formulations of a theory of change, and recognition that systemic feedback is likely to be encountered.

One limitation is that Senge’s work focuses on change efforts within organizations. In such relatively closed settings, simplifying diagrams may be quite appropriate. However, the real-world complexity of community initiatives may not be well represented (Flood 1999 p71). Causal-loops defined at one scale may have no legitimacy at larger or smaller scales.¹³

Moreover, power and resource flows change over time. Pushback may lead to new configurations of power. Archetypes may shift over time. Analysts have noted that causal-loop diagrams cannot easily account for these changes over time (Williams 2005b). Moreover, changes in resource stocks and flows cannot accurately be modeled by causal-loops, limiting their utility for long-term efforts (Richardson 1986).

13 This issue plagues any modeling exercise. Yet CLDs suffer the most of the three reviewed here, since boundaries are not explicitly defined. Both CATWOE and CDE exercises force a modeling exercise to define specific boundaries and then reflect on whether they are adequate to the issue at hand.

The very simplicity of the causal-loop diagram may also force the focus of a food systems effort into too narrow a viewpoint. It may well be that those who, for instance, focus on the growth of CBFS in the US, would overlook the fact that foreign producers, say coffee producers in Costa Rica or soybean growers in Brazil, are affected by decisions made in US consumer markets, and vice versa. In this case, what is viewed by “growth” by local practitioners may be seen as “push-back” by, say, Chilean grape growers. If these distant stakeholders are not at the table, their interests may be overlooked.

Further, Flood argues that consensus-building itself may be inappropriate, since various stakeholders may hold inherently different interests (Flood 1999 p71). Enforcing a consensus may overlook or obscure the power dynamics within the effort itself. Consensus may be artificially imposed by the most powerful or vocal voices at the table, which will skew the modeling process (Flood 1999 p70).

Although causal-loop diagrams are *systemic* in illustrating feedback effects, they may not be adequate to modeling a system over time. This approach may be most useful for people in early stages of systems understanding, who do not yet realize that the system will resist efforts to change it, or with groups that have not had occasion to make their own assumptions about feedback loops explicit. It may not resonate as well among those who have already experienced systems push-back in their own lives.

Nor do causal-loop diagrams appear to lend themselves well to more quantitative assessment, being primarily qualitative images of action and feedback. If one were to model the US farm economy at various points in time, and correlate this to the chart of farm credit diagram above, distinct causal-loops could be applied to different eras of farm production and diverse policy regimes. The data would stand independently of which causal-loop image might be selected for any given era. Still, in Figure 1 (see page 145), knowing the extractive nature of the mainstream commodity economy adds a new level of understanding to the balancing force that limits the growth of CBFS. Thus it suggests caution for those evaluators who might recommend integrating community systems more closely into the mainstream.

ONE “SOFT SYSTEMS” APPROACH: CATWOE ANALYSIS

The issue of multiple perspectives that plagued Senge’s approach to framing consensus is addressed directly by Soft Systems Methodologies, which place strong emphasis on the importance of multiple viewpoints, and upon making explicit the difference between the “reality” we experience, and the models we create as we work in varied contexts. CATWOE is one such method.

Acknowledging that our models of systems are always simplifications, and seldom conform exactly to “real-world” contexts, SSM works diligently to develop models that help explain that reality. Yet we are always limited by our own models. Thus we often test our own *understandings* of systems, more than the systems themselves. Since most people interpret the systems in which they dwell through

their own first-hand experience, comparing these experiences is key (McKinney 2002).

To many practitioners, the first Soft Systems step is to define the “problem situation” at hand. Right away, I would suggest that use of the word “problem” be avoided.¹⁴ This can immediately cast a shadow of negativity or powerlessness over the discussion at hand. In my professional experience, it is more effective to use the word “issue” rather than “problem,” since this leaves the doors open for more comprehensive, and less negative analysis. “Problems” may indeed be opportunities.

I am told that Checkland came to realize this in his later work (Williams 2005b). Flood offers a deeper analysis, stating that many so-called “problem situations” are ongoing, rather than discrete. Modeling them as a moment in time, or as a single problem to be resolved, will not suffice. Rather he suggests a focus on “interacting issues and dilemmas to be managed on a continuous basis” (Flood (1999)). This approach lends itself quite well to situations where power interests may persistently conflict, and to contexts in which a recurrent dilemma will not be solved, but must be addressed.

Preceding a “CATWOE” analysis, many practitioners begin by asking participants to draw “rich picture” diagrams – unstructured images that participants create to illustrate key system dynamics they experience in their contexts. Core forces or “interacting issues” are identified from these diagrams, or from subsequent discussion. From this rich picture, one or more singular perspectives (*holons*) are selected. For each specific perspective, a proposed transformation, or cluster of key elements (using the “CATWOE” acronym¹⁵) is defined:

- C: the Customers who benefit from this system;
 - A: the Actors who transform system inputs into outputs;
 - T: the Transformations that are made;
 - W: Worldview, the relevant viewpoints and assumptions made by various stakeholders;
 - O: the “Owner” to whom the system is answerable; and
 - E: the Environment that influences but does not control the system.
- (Williams 2005 p6).

Each unique set of CATWOE elements leads to its own “root definition” (or purpose) of the food system under analysis.¹⁶ Multiple CATWOE constructs could be created by focusing one at a time on the viewpoints and interests of each of the diverse stakeholders present.

14 See also Flood (1999) p6, 71 and Chapter 10.

15 Some practitioners argue that a more insightful approach uses a “BATWOVE” analysis; in which “consumers” are separated into “beneficiaries” and “victims.” Furthermore “beneficiaries” and “victims” can be ideas as well as people. Others suggest that it is most useful to begin with T (transformations), W (world view) and O (owner) as initial steps in such analysis.

16 For more on root definitions, see Leonard & Beer (1994).

For example, two separate CATWOE constructs might be defined for a given community foods effort, depending upon the point of view to be taken:

ROOT DEFINITION 1 (*Perspective: increase efficiency of food production*)

Customers: Food consumers who buy directly from farms through buying clubs, farmers markets, coops, or community-supported agriculture (CSA) arrangements.

Actors: Food producers who seek to meet this consumer demand.

Transformations: Potential investments to increase the efficiency, scope, or size of these food-producing firms.

Worldview: For this example, we will assume that food producers value commercial efficiency highly. (Of course a variety of other potential worldviews are also possible, including those which place greater value on community connections, on organic food, or those of farmers who wish to place a limit on their workday, etc.)

Owner: The landowner or the owner of any given food business being considered.

Environment: A vibrant local discussion of getting healthy local foods to the region's residents.

The "root definition" (or purpose) of the food system described above would be to increase the efficiency of food production in a given community.

ROOT DEFINITION 2 (*Perspective: build local infrastructure*)

Customers: Civic leaders who wish to assure a steady supply of food for local residents.

Actors: Small groups of citizens already engaged in diverse healthy foods activity.

Transformations: Potential infrastructure investments (ie communications, finance, and facilities) that connect these disparate efforts into a more effective, more highly linked system.

Worldview: An assumption that the stronger the prevailing community connections, the greater will be food security.

Owner: Residents of the region.

Environment: Policy discussions toward local food security.

The purpose defined in this second example would be to increase the region's food security by connecting various food-related initiatives into a more self-conscious system.

In a classic soft systems methodology the next step will be to develop a visual model based on each CATWOE and according to a specific set of systems principles (Williams 2005). However, there are many variations. For instance, CATWOE configurations may inspire revised "rich picture" diagrams, or lead to revised CATWOE definitions (new root definitions), each new step potentially adding new levels of understanding. In such an iterative process, participants may

generate new systems insights by working together – even when self-interests are not identical.

Diverse viewpoints would, for example, shape interpretation of data such as that found on Chart 1 (page 6). The decline of local credit sources would likely be seen as a negative development by local lenders, or by those who care about building community capacity. To a commercial lender, of course, this same trend may be considered desirable.

As can be seen from these two simple examples, the same local food system can appear quite different from diverse viewpoints. How the prevailing local food system is understood, how it is portrayed in diagrams, and how it is modeled in creating a theory of change will depend on the selection of CATWOE elements. Evaluators might well work with local participants to show how these differing “root definitions” emerge out of different constructs, and in turn lead to different evaluative assessments of “worth.” By examining the diverse challenges that emerge when the differing views of varied stakeholders are adopted, hopefully a more integrated understanding of the complexity of the local systems would result.

Either of these “root definitions” would be enhanced by an awareness that powerful economic structures extract wealth from any given rural locale. Evaluation under root definition one, for example, would be made more complete by realizing the forces that act against the growth of any individual farm or processor that may wish to respond to local food demand. Under the second root definition, investments in *networks*, rather than in specific *firms*, are favored. Either insight might shift the focus of evaluation.

This method does seem to add important depth, when compared to causal-loop diagrams. CATWOE creates a structure that can be embraced by beginners. This approach also has strong heuristic value. Change over time can be accommodated here by tracking transformations that may alter the system. Less clearly-bounded situations can be modeled iteratively using successive CATWOE definitions.

Many evaluators have seen SSM as an effective tool to use with groups of collaborators in learning and making meaning together (Williams 2005 p2). Leonard and Beer (1994) argue that SSM is best used when there is uncertainty about the issues to be confronted, at a point when a group individuals might lay down their organizational perspectives to define new approaches. SSM may not be appropriate, they argue, when an issue is already clear. They suggest that SSM may lend credibility to efforts to interpret hard data, since interpretation of such data is subjective. The clarification of diverse viewpoints may lend rationality to this interpretation, especially when the viewpoints of those who have been marginalized emerge. (Leonard and Beer 1994 pp 37, 32, 34).

CDE ANALYSIS – ONE COMPLEX ADAPTIVE SYSTEMS APPROACH:

Diverse Complex Adaptive Systems (CAS) approaches, both quantitative and qualitative, focus on the complex and changing nature of systems. CAS approaches

acknowledge that as people within a system take action, the system itself changes, resulting in combinations of structured and less-structured elements. Some recognize that activity in a CAS may be random. Or groups of individual agents may self-organize to create greater stability. Some computer models track how groups of individual actors, following simple rules appropriate to their contexts, may create complex patterns of behavior across the system. Evaluation often draws upon these insights.

Flood considers spontaneous self-organization to be a special form of *emergence*. (Flood 1999 p2) This refers to unexpected patterns of complexity that result from simpler processes, but could not be predicted from the rules followed in the less complex process. Evaluators also look for “*attractors*,” patterns, clusters of energy or resource flows that tend to create stability among disorder, and that may provide the backbone for lasting systems change.¹⁷

Many consider analysis of time-sequence data – measurements taken at regular intervals, using similar techniques, that show historic trends or patterns – as the most useful way to illuminate changing resource flows or emerging attractors (Eoyang and Berkas 1998 p14).

In this paper, one specific qualitative approach will be used: CDE (Eoyang and Berkas 1998 & Eoyang 2004). This method raises three core questions:

C: what is the Container in which we work – what boundaries do we place around our context?

D: what are the “Differences that make a difference”? What issues or dilemmas do we find posed within the system at hand, or what measures can we identify, that tell us whether systems change is proceeding in the proper direction?

E: What are the Exchanges that occur within and across system boundaries? How might these exchanges be altered or shaped to promote desired changes?

For our discussion, looking at the big picture, the Container is the US farm and food economy, as viewed in its global context. “Differences that make a difference” might include the competing views of diverse participants in the farm and food economy, or the differing social connections formed by people who work to build community affinity, rather than purely commercial relations. The strength of local credit sources, as discussed above, would be one such measure of the Exchanges that occur. Obviously, many alternate CDEs could also be defined at different scales.

In CDE, the different realms of order and disorder, structure and randomness, self-organization versus imposed, are all embraced. Simply by assuming a high degree of disorder, CDE poses processes that are not dependent upon the boundaries of a specific firm or department. There is a very immediate sense of *change over time* built into the approach, and a humble sense of the limits of human capacity to intervene amidst the complexity and inertia of prevailing systems.

17 Definition of emergence adapted from Flood, 2. Definition of attractors adapted from Eoyang.

CDE lends itself quite well for evaluations involving those internal to a process, less susceptible to manipulation from external parties. The need for multiple perspectives seems to emerge easily from its complex understanding of randomness and structure. The open-ended nature of CDE creates strong heuristic value.

The caution to be made here is that, having defined an inner tension, diverse stakeholders may find they don't hold common interests, and do not agree. This may not be a path to greater unity, yet even making differing views explicit can strengthen future collaboration, if the evaluator is able to present a positive approach.

CDE may be less useful in highly structured situations. Furthermore, although it appears to be a powerful tool for those within a complex environment to use to understand the complexities they experience, that very immersion may make it difficult for participants to achieve the detachment of "real world" and "model" that SSM strives toward.

CONSULTING "WISE PRACTITIONERS"

No matter which modeling method is pursued, consulting with "wise practitioners" is often useful. These are source people, well-immersed in the context to be evaluated, who have gained special insight, or who are especially articulate in taking a broader view of the issues at hand. One way to think about them is that they are the people who have their feet planted firmly in the mud of the situation, but also keep their eyes open, reading the winds. They draw, therefore, from at least two perspectives at once. They know intimately the issues that are faced on the ground, and yet they can speak about this insight from a position that is broader than their personal self-interest.

One of the key challenges, of course, is to know how to select the appropriate wise practitioners. Often the most visible or popular voice is not the most informed, or the most informative.

Many experts recommend that a group of specialists from a variety of disciplines be brought in (Flood 1998 p68), so that the blind spots inherent to any specialty will be less persuasive. This can be a highly effective and stimulating process. Yet Flood, referencing D T Suzuki, offers a strong caution here, pointing out that if to specialize is to break reality down to its component parts, this may be like breaking a mirror into pieces and reassembling it. Once put back together, an image may appear in the mirror, but it will be a broken image – not as whole as if the mirror had not been broken in the first place. (Flood 1998 p14)

There is strong alternative path, less easy to define or standardize, that can also be effective if participants are chosen carefully. This is to select one or more people who are themselves generalists – people who never occupied, or who have already stepped outside the confines of, any given specialization.

A special case of this is in a community setting. Rooted community residents tend to be generalists, since they have experienced the systemic pressures that apply to their lives intimately, and over time. If a group of such immersed residents

has worked well together, gaining mutual trust and respect, they are exceptionally well positioned to provide systemic insights. These are people who, to adapt Flood's words, have learned inside feedback structures. (Flood 1998 pp68-69) They have also engaged in thoughtful reflection about their own mental frameworks (Flood 1998 p68). They are able to "see the world through the eyes of another."¹⁸ They know how systems push back, and are less likely to rely on linear models than specialists would.

Such sources may have a visceral, rather than a modeled, insight into systems dynamics. People in poor or marginalized communities, for example, often have a far richer understanding of the powerful than do the powerful themselves – since they have been forced to deal with the consequences of decisions made by the powerful. The reverse understanding is rare.

For an outsider who interviews such an immersed practitioner is the need to gain enough mutual respect with the interviewee that one will be trusted with the best stories, rather than the official tales that are often told well-meaning outsiders. It is also essential to have enough first-hand experience in the context itself, or enough allies who do, that wise insights can be separated from dishonest or self-serving statements.

Conclusion

The complexity of systems thinking creates a distance for many lay people. Yet these same lay people may have strong intuitive understandings of the systems in which they work. Especially in the process of constructing, evaluating or revising theories of change, the modeling methods outlined here hold great utility by providing a common language to the discussion.

Each of the three modeling methods discussed in this paper hold utility for different constituencies and varying situations. Each should be considered valuable tools for the evaluator's toolbox. It may be fruitful to blend their use in many settings, guided by the cautions raised in the evaluation literature as outlined here.

In each case, incorporating the testimony of wise practitioners as well as insights from quantitative data, helped to deepen use of the evaluation tool. While closed-loop diagrams may be too qualitative to be well-suited to the incorporation of data, new insights were gained by consulting well-chosen data sets. CATWOE definitions were informed by data that illuminated the experiences of various stakeholders who experience extractive relationships in different ways. CDE analysis relies heavily, as does much of CAS evaluation, on time-series data. In each case, quantitative analysis was strengthened by incorporating the insights of wise practitioners.

¹⁸

Flood, 63, citing the work of C. West Churchman.

The following table summarizes the qualities of the methods reviewed:

Characteristic	Causal-loop Diagrams (Systems Dynamics)	CATWOE (Soft Systems Methodology)	CDE (Complex Adaptive Systems)
Easily understandable	Yes	Yes	Yes, but CAS difficult
Expresses feedback and other systemic qualities	Yes	Yes	Yes
Heuristic value (leads to future learning) (Forrester 1985) ¹	Yes	Yes	Yes
Expresses separation between “reality” and “model”	Yes	Yes	Yes
Lends itself well to lay use	Yes	Yes	Perhaps
Builds agreement among diverse stakeholders	Yes	Yes	Yes
Designed to build consensus among participants	Yes	No	No
Expresses change over time	No	Perhaps	Yes
Embraces multiple perspectives	No	Yes	Yes
Expresses power dynamics at work within the initiative itself	No	Perhaps	Likely
Designed for use in a highly bounded, stable or organizational context	Yes	Perhaps	Seldom
Expresses stocks (accumulation) and flows (movement) of resources	No	Perhaps	Yes
Lends naturally to measurement of key dynamics	No	Perhaps	Yes

References

- Crawford, A E, Nettle, R A, Armstrong, D P, and Paine, M S. 2002. *A framework for aligning social and technical orientations to farming systems research, development and extension – an Australasian experience*, 6. Viewed July 1, 2005, at conference.ifas.ufl.edu/ifsa/papers/c/c4.doc.
- Eoyang, Glenda H and Berkas, Thomas H. 1998. *Evaluation in a Complex Adaptive System*. Available at <http://www.winternet.com/ffieoyang/EvalinCAS.pdf>
- Eoyang, Glenda H. 2004. *Complex Adaptive Systems (CAS)*. Kellogg Foundation, May, available at <http://users.actrix.co.nz/bobwill/Resources/cas.doc>.
- Eoyang, Glenda. 2005. Personal communication, August 3.
- Evolutionary Agroecology and Biocomplexity Initiative. *EABC Statement of Terms: The Limits of Simplicity in the Agriculture Enterprise: The Genomic Ecosystem-Agroecosystem Gap*. Viewed July 1, 2005 at <http://www.agron.iastate.edu/ffiweds/EvoAgroBioC%20InIt/EvoAgroBioC%20TERMS.htm>.
- Flood, Robert Louis. 1999. *Rethinking the Fifth Discipline: Learning within the Unknowable*. Routledge. 71.

- Forrester, J W. 1985. 'The' Model Versus a Modeling 'Process'. *System Dynamics Review* 1(1): 133–134. Summarized by Sastry, M Anjali and Sternman, John D. 1992. *Desert Island Dynamics: An Annotated Survey of the Essential System Dynamics Literature*. MIT Sloan School of Management System Dynamics Group. Viewed at <http://web.mit.edu/jsternman/www/DID.html> on August 31, 2005.
- Friedman, Milton, and Schwartz, Anna. 1963. *A Monetary History of the United States, 1867–1960*. Princeton University Press
- Galbraith, John Kenneth. 1954. *The Great Crash*. Boston: Houghton Mifflin. "Had the economy been fundamentally sound in 1929, the effect of the stock market crash might have been small."
- Latham, A J H. 1981. *The Depression and the Developing World, 1914–1939*, London: Croom Helm.
- Leonard, Allenna and Beer, Stafford. 1994. *The Systems Perspective: Methods and Models for the Future*. In, *Future Research Methodology*, 33. American Council for the United Nations University Millennium Project. Available at www.futurovenezuela.org/_curso/6-sysmeth.pdf.
- Lichtenstein, Benyamin Bergmann 2000. *The Matrix of Complexity: A Multi-Disciplinary Approach for Studying Emergence in Coevolution*, Version 9.3, October. Note that the Evolutionary Agroecology and Biocomplexity Initiative at Iowa State University defines agriculture as a complex adaptive system.
- McKinney, Earl H, Jr. 2002. *Leveraging the Hidden Order of Systems*, 5. Viewed July 1, 2005, at <http://www.bauer.uh.edu/parks/fis/systemsMckinney.htm>.
- Meter, Ken. 1990. *Money with Roots*. Minneapolis: Crossroads Resource Center, available at <http://www.crcworks.org/roots.pdf>. See also Meter, Ken 1983. *Green Isle: Feeding the World and Farming for the Banker*. Minneapolis: Crossroads Resource Center and Farmer Labor Education Committee, 3. Available at <http://www.crcworks.org/gi.pdf>.
- Meter, Ken. 2003. *Food with the Farmer's Face on it: Emerging Community-Based Food Systems*. Media guide published by W K Kellogg Foundation. Available at <http://www.wkkfweb.org/FSRDFullGuide.pdf>.
- Perkins, Van L. 1969. *Crisis in Agriculture*. Berkeley: University of California Press
- Richardson, G P. 1986. Problems with Causal-loop Diagrams. *System Dynamics Review* 2(2): 158–170. Summarized by Sastry, M Anjali and Sternman, John D (1992). *Desert Island Dynamics: An Annotated Survey of the Essential System Dynamics Literature*. MIT Sloan School of Management System Dynamics Group. Viewed at <http://web.mit.edu/jsternman/www/DID.html> on August 31, 2005.
- Rothermund, Dietmar. 1996. *The Global Impact of the Great Depression, 1929–1939*. London: Routledge.
- Senge, Peter. 1990. *The Fifth Discipline: The Art and Practice of The Learning Organization*. Currency Doubleday.
- Sørensen, Lene and Vidal, René Victor Valqui. 2002. The Anatomy of Soft System Approaches. *Economics Analysis Working Papers* 1:8. Viewed July 1, 2005, at <http://eawp.economistascoruna.org/archives/vol1n8/>.
- Temin, Peter. 1976. *Did Monetary Factors Cause the Great Depression?* New York: Norton.
- Timoshenko, Vladimir. 1933. *World Agriculture and the Great Depression*. Ann Arbor: University of Michigan Business Studies, Volume 5, Number 5 (School of Business Administration, Bureau of Business Research).
- USDA/NASS Agricultural Census 2002. Viewed July 13, 2005, at http://www.nass.usda.gov/census/censuso2/volumer/us/st99_1_002_002.pdf.

- Wee, Herman van der. 1972. *The Great Depression Revisited*. den Haag, Netherlands: Martinus Nijhoff.
- Williams, Bob 2002. *Evaluation and Systems Thinking*. Available from <http://users.actrix.co.nz/bobwill/>
- Williams, Bob. 2005a. *Soft Systems Methodology*. Available from <http://users.actrix.co.nz/bobwill/Resources/ssm.doc>.
- Williams, Bob. 2005b. Personal communication, August 4.

USING SYSTEMS CONCEPTS IN EVALUATION

Systemic Evaluation in the Field of Regional Development

Richard Hummelbrunner

Richard's chapter makes explicit what is implied in other chapters. Evaluation can be considered as a system itself. Applying systems concepts to the evaluation task can be as insightful as applying systems concepts to the situation that an evaluator is evaluating. Making that leap of reframing what evaluation is then allows you to draw on new evaluation methods. Richard describes a range of systemic tools drawn from his experience evaluating regional development in Central Europe. They are primarily used to address stakeholder differences based on exploring relationships and perspectives – something familiar to all evaluators.

Introduction

Regional (and local) development processes are characterised by the following qualities:

- *Openness*: Regional development is increasingly dealing with open tasks, whose results cannot be known in advance (eg improving competitiveness, promoting innovation). In this case, only general objectives and processes can be defined beforehand, but concrete solutions and appropriate approaches will emerge only gradually.
- *Linkages*: The success of regional development policy depends on the interaction of economic, social, cultural, and physical resources within a territorial unit and on the quality of collaboration between key actors having access to or being responsible for these resources.
- *Unpredictability*: The key players in regional development processes (providers as well as recipients of support measures) are social actors (institutions, individuals). The behaviour of these social systems is not “linear” and predictable, but subject to unforeseeable changes.

As a consequence, regional development should be seen as an open process of transformation, which continuously needs to be shaped and where it is necessary to react timely to changing conditions or new opportunities. Evaluations can make an important contribution for such ongoing adaptation, provided they are oriented towards strengthening the management and reflective capacity of the involved actors. One of the main challenges is to deal appropriately with complex social systems and the interaction of a wide (and often changing) variety of actors with different values, interests, and motives.

Evaluation findings often reveal a rather diverse picture of a programme/project, particularly when viewed through the eyes of various stakeholders. However, unbalanced or simplistic attempts to reduce this complex picture will

not only harm the credibility of an evaluation, but also bring forth resistance from those who do not feel themselves properly represented. In particular, evaluations should go beyond merely illustrating differences in opinions (eg by visualising diversity through rating, ranking or mappings) and effectively work with them.

Systems thinking can be of great help to avoid undue simplifications and provide useful tools for dealing practically with differences. It can contribute looking beneath the surface of observable phenomena and to identify underlying patterns and causes. By revealing core dynamics it can help to provide simple – not simplistic – insights into the functioning of complex systems. And systems approaches can also contribute to improve the use of evaluations. Because generating new insights and improving joint understanding of issues across a range of stakeholders is an important pre-requisite for sustainable learning effects.

The present paper starts out by explaining what are the basic elements of my systemic evaluation approach. Section 1 takes a look at evaluation as a system and as a regulatory mechanism. And it explains the theoretical implications and practical consequences of two essential ingredients of my personal stance: regarding evaluation as an intervention in the evaluated (client) system which is aimed at increasing the client's capacity to understand, solve problems and change. And dealing differently – and more appropriately – with differences as the visible part of internal regulatory mechanisms of open (social) systems. In this respect section 2 outlines some selected tools which I use for dealing with stakeholder differences and which originate from various fields of “systemic practice” (ie family therapy, management consultancy, organisational development). Section 3 contains some concluding remarks concerning the utility of systemic evaluations and summarises their basic requirements in terms of process and evaluator roles.

A Systems Approach to Evaluation

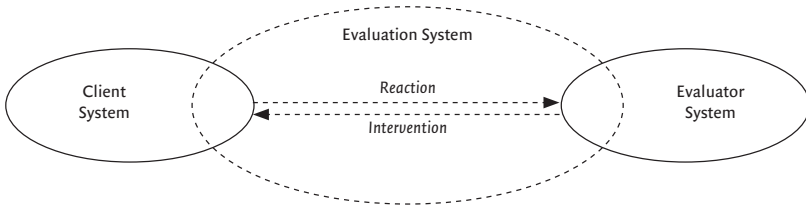
1.1 EVALUATION AS A SYSTEM

Evaluation can be regarded as a specific case of observation, which takes place in a joint system established between two main partners:

- *Client System*: Consists of the funders, the operators (ie managers of the programme/project to be evaluated) and the concerned public (other stakeholders such as beneficiaries, partners, additional intended users of evaluation results)
- *Evaluator System*: The experts which are commissioned to undertake the evaluation

The Evaluation System (figure 1) is usually established by contract and is limited in time. It has a joint focus based on the evaluation purpose, and a structure to serve it. Elements of this structure are nodes of communication (ie meetings with

Figure 1: The Evaluation System



fundlers, steering group, synthesis workshops) and their respective linkages as defined in the design of the evaluation process (eg work packages, activities).

The Evaluation System incorporates elements of the two constituting partner systems and the illustration above can be used in establishing the respective boundaries: For instance, which elements of the client system take part in specific nodes of communication (eg steering group, workshops)? Who from the client system participates in – or even carries out – evaluation activities (eg self-assessments, surveys). An important question to be clarified in any evaluation assignment is whether – and to which extent – the funders of the programme/project are to be included in the evaluation. From a systems perspective, they are part of the Evaluation System, whether they like it or not. But they often prefer to be treated in a detached manner, separated as much as possible from the programme/project being evaluated. However, in my experience, looking at the evaluation as a system can effectively critique the tight boundary which they prefer to draw around them – and lead to their constructive involvement despite initial perturbations.

All three systems are part of each others environment. Whatever happens in the Evaluation System can affect the constituting systems, and vice versa. And evaluation is an intervention in the Client System, bringing forth reactions in the latter which again can have an influence on the Evaluator and Evaluation Systems and so forth. It is a circular process, by which all three systems mutually influence each other. Intervention means to apply external influence upon a system with the aim of inducing change. But as Simon (1997) explains, social systems are self-determined and can only change themselves, this cannot be done by an intervener, no matter what resources, power etc. are applied (at least not in the long run and in a sustainable manner). Ultimately every system decides on its own – and according to its own logic.

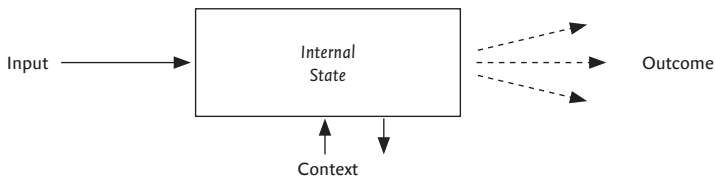
Due to these pre-conditions, interventions in social systems cannot be linear or directive, their outcome is uncertain and they bear certain risks (Willke, 1995). In order to limit these risks and increase the chance for success, some precautions should be taken with interventions (Königswieser and Exner 1998):

- *Beware of connectivity:* To be accepted, an intervener should have a profound understanding of the system's internal structure and act accordingly. This

requires, eg to use similar language, respect dominant rules and behaviour patterns, build on existing concepts and values, address topics conceived as being relevant by the system.

- *Keep interventions balanced:* Expose to external views in a moderate way, do not confront too openly. Do not only seek change, but also make aware what should be maintained. Balance contradictions and ambivalent tendencies (eg the good in the bad).
- *Base interventions on prior hypothesis:* Conceive interventions as a circular process. Before intervening, collect information about the system, formulate hypotheses about the situation and the intended effect(s) of the intervention. After the intervention, collect information about it's effects, reformulate hypothesis and so forth.

“Systemic” interventions are therefore targeted forms of communication between (social) systems, which respect the autonomy of these very systems. Although they are conceived by an external intervener, they should be designed in the terms of the system which the intervention is aimed at. And interventions can only stimulate a system’s self-steering mechanisms, they are not aimed at intended effects in a linear way. It is the structure of the “target” system which is decisive for the success of an intervention. But according to v. Foerster (1970) social systems are “non-trivial machines”, they can react differently at different times to the same input (or intervention) – depending on their internal state. Their behaviour (outcome) is not linear, it can neither be explained from inputs nor their internal states, but results from the interaction of both:



As a consequence, systemic evaluation is an intervention in a Client System with the essential aim of modifying the client’s internal state in order to improve the chances for producing desired outcomes in a sustainable manner. This essentially means to increase the client’s capacity to understand the situation at hand, solve occurring problems and change in a way that contributes to expected solutions. And at a more operational level, evaluators should take these complex relations and linkages into account when planning or implementing their interventions during the evaluation process (eg interview sessions, focus groups, surveys, presentation of findings).

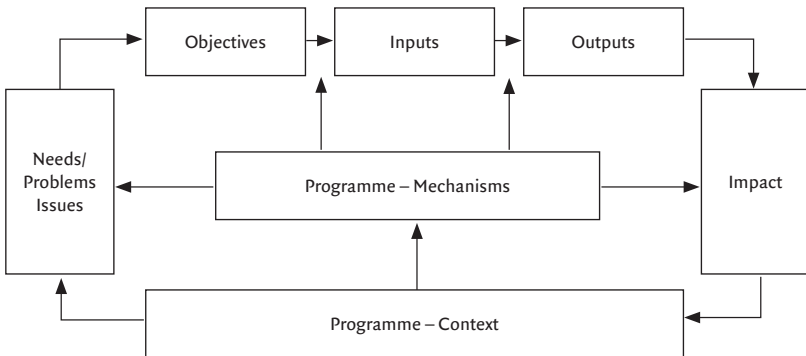
But viewing evaluations as a system also has important implications for

defining their content, notably when determining an evaluation's coverage, scope and level. From a systemic perspective two principles should be borne in mind in this respect:

- The programme/project to be evaluated (evaluand) should be structured as a system, ie outlining the essential elements and their relations.
- The unit of observation should be the evaluand *and* its environment, ie what lies beyond the boundaries of the evaluand when seen as a system.

In the case of regional development programmes¹, their basic elements consist of objectives, inputs and (expected) effects, ie outputs and impacts. The main relations are the mechanisms which link these elements: for example, planning and decision making mechanisms that determine how inputs (eg financial, human, physical resources) are applied in order to achieve objectives. Or implementing mechanisms (activities, management arrangements) which are foreseen to transform inputs into outputs. In addition, programmes take place in an operational context, which influences implementation in manifold (and often unforeseeable) ways.

Thus regional development programmes can be structured as systems, by extending logic (or change) models to a systemic evaluation framework in the following manner:



The basic elements of programmes are objectives, inputs, outputs and impact (which might be further differentiated as immediate impact, outcomes etc). Their links are influenced by the a set of mechanisms which are foreseen in the Programme:

- transformation of objectives to inputs via decisions on resources (eg financial resources, funding conditions, human resources)

¹ These programmes are characterised by multiple objectives and domains of intervention. They normally consist of a set of support measures with specific objectives and budget. These measures are implemented via operations or projects within a given time frame and pre-defined funding conditions.

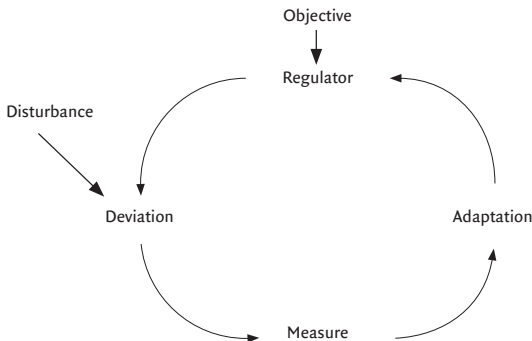
- transformation of inputs to outputs via communication with beneficiaries, project generation and project selection procedures/criteria
- transformation of outputs to impact through the use of outputs (= projects) by project owners, target groups or implementing partners.

And the programme context (eg physical conditions, institutional or legal framework conditions) influences the way in which programme mechanisms are applied, and can in turn be influenced by programme impacts. Therefore the programme elements are linked to mechanisms and context in a recursive logic. And the achievement of effects (ie outputs, impacts) is not seen in an isolated manner, but takes the actual functioning of the programme or relevant context conditions into account. Impacts are conceived as the result of specific mechanisms acting in a specific context, linked by feedback loops. Because impacts modify the context, this has potential effects on programme mechanisms, which in turn can affect the transformation of inputs into outputs and impacts, and so forth². For instance, training activities or the establishment of facilities in the early stage of a programme changes initial context conditions and influences the way in which activities can be carried out at a later stage.

By systematically making a clear distinction between internal (=mechanisms) and external factors (=context), evaluations can lead to learning effects of actors involved in implementation, because they are provided with information on how to modify their mechanisms in order to improve the achievement of effects.

1.2 EVALUATION AS A REGULATORY MECHANISM

In cybernetic terms³, evaluation is essentially regarded as an external regulator during the programme cycle (planning – implementation – monitoring/evaluation):



- 2 These recursive links are the main difference with regard to the “Realist Evaluation” approach, which only foresees linear relations between context and mechanisms (Context – Mechanism – Output configurations). See Pawson & Tilley (1998).
- 3 See Dale Fitch’s chapter for a full explanation of what this means.

An objective is set and a measure (eg indicator) established, which allows to observe whether actions achieve the objective. When this measure shows a deviation (due to internal or external factors), evaluation functions as a regulator and proposes adaptations with the intention of modifying the action so the original objective can be met.

Thus evaluation constitutes a negative feed-back loop whereby differences from a desired state are counteracted by actions in the opposite direction (ie if the value of the measure is too low, then actions are taken to increase it). Differences from original targets are a priori regarded as negative and in need of corrective actions to put a programme / project back “on track”. And their success can be assessed by comparing outcomes against initially defined objectives.

Such an approach is suited (and was originally conceived) for closed non-living systems, which operate in a closed environment and whose elements do not possess internal dynamics. It is modeled according to regulatory processes in the physical or mechanical world (eg thermostats). Under such conditions, fixed targets can be established and constant “linear” relations between actions and their results maintained.

Territories, however, are open living systems, they are in continual exchange with their environment and their elements/subsystems can change over time. The various subsystems (eg political, administrative, economic, social) interact within a given territory and the development of one system might cause adaptation processes of another system (and vice versa). Programmes / projects are implemented in this environment of interacting social systems and – due to their non-linear behaviour – rarely act one-way, but might also trigger processes, which can neither be foreseen nor reduced to original plans or intentions.

The resulting non-linear behaviour of open systems has two important implications:

- *Relationships between cause and effect are neither proportional nor transparent:* Every action can be both – cause and effect; therefore “linear” cause-effect links are replaced by “circular” interaction patterns, which are made up by (negative and positive) feed-back loops and regulate the behaviour of a system. But it is difficult to trace all the linkages and effects, therefore social systems can never be thoroughly analysed nor understood.
- *Changes are essentially self-organised:* Open systems develop their own internal mechanisms of regulation and stabilisation (autopoiesis) and cannot be controlled externally – at least not in a direct mechanistic sense (Maturana and Varela 1992, Probst 1987). Because changes in the environment are ambivalent: they can be disturbing and trigger corrective as well as defensive action – or they are the source for further development, leading to modifications in relations as well as new (inter)actions. Adaptation processes in open social systems follow internal mechanisms rather than external influence (eg recommendations of evaluators).

Open systems are stabilised through constant renewal of their elements – “everything changes unless someone /-thing ensures that things remain as they are” (Simon 1997). Therefore differences from original states are an inherent feature of open systems in order to assure their stability. And changes in short term targets or plans are often necessary for the achievement of long-term objectives.

As outlined in section 1, regional development programmes can be regarded as open systems. And one should be aware of these fundamental characteristics when evaluating them. Instead of conceiving evaluations as an external regulation (which is appropriate for closed systems) they should be designed in line with the internal regulatory mechanisms of open systems. Above all, this requires a different attitude towards change (eg differences from plans), regarding them as visible expression of internal regulation:

- Analysing differences in output (as well as results and impacts) can help to assess the appropriateness of a programme in view of the given environment (eg framework conditions, needs of target groups, interests of implementing partners). But it can also provide valuable indications about the internal dynamics and self-organising forces which are at work within social systems and thus improve the understanding of the operating environment.
- For evaluations to work this way they should not be limited to observing intended effects and routes, but instead look at the entire range of effects triggered by the programme, irrespective of whether they are in line with original intentions. Exceptions, discontinuities, unexpected results and side effects are valuable sources of information on the programme being evaluated. They can provide useful clues eg for relevant internal/external changes, newly emerging challenges, innovative or “informal” ways of handling situations, which can help to improve implementation.

If this is not taken into account, evaluations risk counteracting these internal mechanisms, which might result in misleading conclusions and even counterproductive recommendations. Insisting on the implementation of original plans despite relevant changes in the operative environment might have severe counter-intentional effects and ultimately result in failing to achieve objectives. Evaluations should therefore progress from adaptive, single loop learning (“doing things right”) to generative double loop learning, which opens the view to other alternatives (“doing the right things”). But this requires to change the internal state of social systems, notably their relevant patterns of cognition or behaviour and to arrive at a different – and often more comprehensive – understanding of the generative mechanisms at work.

This need not necessarily be a cumbersome and time-consuming exercise, because the manifold observations of stakeholders can already provide a valuable source of information – provided they are noticed and put to use. Some practical approaches in this direction are to:

- reflect with stakeholders on relevant (internal/external) changes during implementation and on whether and how the implementing system has reacted to these changes.
- focus on meaningful differences during implementation, either by periodically monitoring change (instead or in addition to monitoring indicators) ⁴ or by documenting changes in the behaviour of implementing partners, which are relevant for achieving outputs and results ⁵.
- systematically compare planned and factual states (eg on targets, indicators) and attribute explanatory factors of these differences either to internal or external mechanisms (by using the systemic evaluation framework described in section 1.1).

1.3 DEALING APPROPRIATELY WITH STAKEHOLDER DIFFERENCES

Another important set of differences which need to be dealt with in evaluating regional policy concern opinions and views among various stakeholders. Processes involving social systems are very much based on observation, and (inter)actions are the results of actions observed with others – and the meaning attributed to them. Each actor constructs an internal mental map for orientation, which serves as a frame of reference for action or choosing among alternatives (Bandler and Grinder 1975).

Thus the existence of diverse and often conflicting views of different individuals or groups should be seen as the rule – and not as an exception from the ideal of one single truth. And these differences cannot be solved by giving preference to one particular view or by synthesising them through an objective judgement (eg by an evaluator). Because ultimately everyone one is right – but only within the boundaries of the respective mental maps (Simon 1997)!

But differences among stakeholders could be treated as a resource instead of an obstacle. They can improve mutual understanding if mental maps are made explicit and visible for others. And they can contribute to a more complete picture of reality by linking individual mental maps and working towards the emergence of collective mental maps (Bandler and Grinder, 1975 and 1976).

Under such circumstance, consensus and mutual understanding should not be taken for granted, but be regarded as temporary compromise despite divergent views and objectives. And this can most likely be achieved if the various actors confront their mental maps, become aware their own limitations and more receptive for the views of others. Thus they gain adequate understanding of the internal mechanisms of perception and of the relevance criteria of the participating systems. And they learn to see things not only from their own point of view but also from that of the others they are interacting (Watzlawick et al 1974).

However, achieving consensus in evaluations is often not possible. But learning can also consist in improved understanding of – and a change in attitude

4 A useful and innovative tool for this purpose is “Most Significant Changes Monitoring“ (Davies, 1998).

5 This is the approach used in “Outcome Mapping“ (Earl/Carden/Smutylo, 2001).

towards – stakeholder differences. To accept differences as part of a complex reality, where the picture of the whole can only emerge by viewing it from multiple angles, and by using or adding different descriptions or explanations in a joint dialogue aimed at reconstructing reality.

Mutual understanding and successful communication between different groups can be facilitated if the following aspects are acknowledged and explicitly taken into account:

- the existence of *different environments* as well as different professional, institutional and cultural framework conditions at local, regional, national and supranational levels
- the existence of a *specific* (and necessarily selective) *repertoire* of what a system regards as meaningful communication (concepts, language, etc), which media of communication (descriptions, numbers, maps etc) are preferred and which representations of “reality” are therefore regarded as adequate
- the principal limits to *cognition* for all social systems (including, of course, oneself) due to the selection processes of self-referential systems (Watzlawick et al 1974). Each one only sees a specific part, but is on the other hand not aware of the blind spots – to quote Simon (1997): “we don’t see that we don’t see”.

On the basis of such an attitude of mutual interest and acceptance the confrontation between social systems and their different views – which, without that basis, usually provokes misunderstanding, rejection and resistance – can be made productive. Eventually mutual understanding can be deepened and a common basis of communication established.

2. Selected systemic tools

Systemic tools are important instruments to achieve “higher level” learning processes as they allow to actually work with differences in order to increase mutual understanding, achieve consensus or bring forth joint solutions. They are well suited to visualize complex realities (eg causal diagrams, concept mappings, social network analysis), allow to better understand differences (eg reframing and dialogue techniques, large group interventions) or to overcome differences in innovative and often surprising manners (eg constellation work, solution focus, paradox interventions).

This section contains some selected systemic tools, which can be used in dealing with stakeholder differences. Since stakeholders and evaluators alike are usually faced with considerable time constraints, only tools have been included which lend themselves for group work and can be applied in short sessions. Their use can lead to highly effective interventions, as they are capable of “making a difference” for participants even within very short time spans.

CONNECTING DIFFERENCES: CAUSAL-LOOP DIAGRAMS

Causal-Loop Diagrams are based on the concept of “feedback”, a language for representing “circular” relations which was originally developed in cybernetics. System elements are linked by two types of feed-back mechanisms:

- (-) Negative feed-back: interaction works as a limiting factor and leads to a compensation process aimed at closing the gap between a desired and the actual state (stabilisation)
- (+) Positive feed-back: interaction leads to an increase of the previous state in the same direction (growth or decline).

Causal-Loop Diagrams are a relatively easy tool for visualising complex relations, they can be generated on paper, pin walls or computers. Since they facilitate communication on complex issues and can be modified rather easily, they are very appropriate for group work.

Although Causal-Loop Diagrams are a tool developed in the First Wave of systems thinking, they can also be applied in a Third Wave mode for boundary critique, ie exploring boundaries through stakeholder discussions⁶. To this end, stakeholders are first asked to provide explanations for a given situation from their individual points of view. As a next step, these different explanations for the same phenomena can be confronted and the differences between these explanations explored in more depth. Stakeholders are asked to justify their specific boundary choices (why have they chosen certain elements or linkages), but also to question the boundary choices of others. And finally a reflection takes place as to which of these multiple explanations are compatible with each other (ie complimentary or mutually reinforcing) and which of them are antagonistic, which might lead to further question their rationale in terms of power relations, back-up evidence or value judgements.

But connecting individual views can also be used to obtain a more comprehensive picture of reality. Linking different descriptions or explanations can lead to multiple descriptions of the same phenomenon, and being able to see and value emerging “both-and patterns” instead of “either-or” relations can create new insights and open the way for new solutions.⁷

For example, the Causal-Loop Diagram on the following page was produced during the evaluation of a Technical Assistance Fund for private enterprises. Low use of the Fund was evident from data previously collected and identified as one of the core issues. Then explanatory factors for this situation were collected from a series of group interviews with stakeholders (eg beneficiary businesses, consultants providing technical assistance, business institutions). And finally these factors were arranged by the Fund operators (who incidentally were present

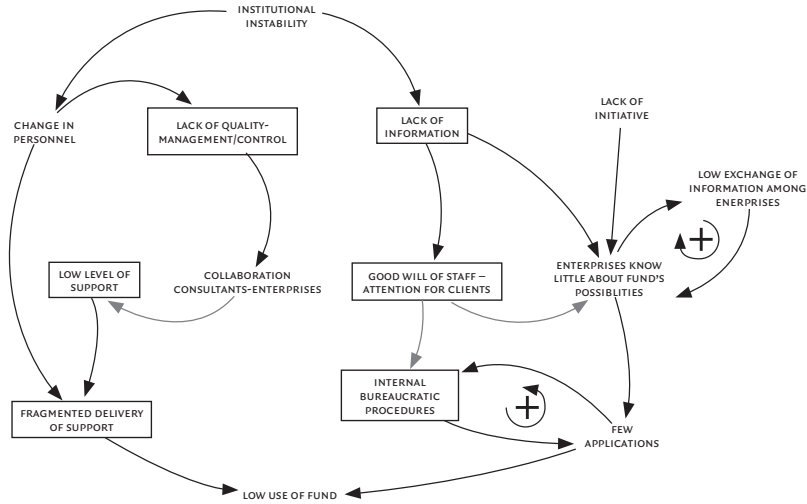
6 See Gerald Midgley's chapter for a full description of the three “waves” of systems thinking.

7 See also the discussion of Soft Systems Methodology in the chapters by Kate Attenborough, Boon Hou Tay, and Ken Meter.

during the interview sessions), and were linked to form reinforcing (+) and balancing (-) feedback relations. This provided Fund operators with a much richer picture of reality and valuable insights from their client stakeholders. And it allowed to identify the leverage points, ie factors which can be directly influenced by the Fund operators and can have considerable influence on other elements. These factors were framed in the Causal-Loop Diagram and actions were designed to change the situation in the desired direction.

In addition to modelling diverse stakeholder opinions, such diagrams can also be used to identify structures underneath observable phenomena (“symptoms”). For instance, when unintended effects of an intervention are linked to the original theory of action (modelled as feed-back processes), their generative mechanisms can be revealed and indications are given on how they could be curbed or even avoided.⁸

Figure 2



USING DIFFERENCES: DIALOGUES

Whereas in discussions or debates standpoints are confronted and – if possible or desired – harmonised with each other, a dialogue is a means to foster collective intelligence. Contrary to a debate, participants do not exchange ideas with the intention of convincing others, therefore it does not make sense to try and “win” a dialogue. Rather it is aimed at producing a collective sense by sharing and exploring various perceptions and insights (Senge et al 1994, Isaacs 1999).

8 Jay Forrest's, Dan Burke's, and Ken Meter's chapters also discuss CLD's. Ken Meter's paper also discusses some of the limitations of causal-loop diagrams.

Dialogue techniques are based on three key assumptions (Bohm 1996):

- *Different observation positions make a difference:* Exposing oneself to different perspectives helps to overcome mental barriers or unilateral thinking and to find solutions and answers which are acceptable for all.
- *A system can only be changed from within:* A dialogue never aims at direct influence or persuasion, but rather at reference experiences which enable the partners to change their “mental maps”.
- *Language counts:* The purposive use of specific language patterns helps to understand the partner’s mental map “from within” and at the same time to overcome its limitations.

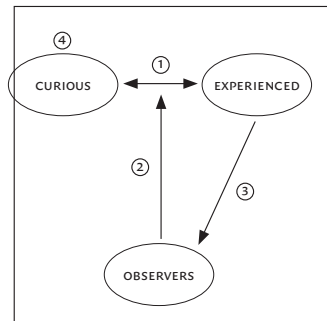
In *Circular Dialogues*, participants have the opportunity to perceive a given theme from at least three perspectives. Guided by facilitators, participants are asked to dialogue in a structured manner, mutually interviewing and observing each other without direct discussions. The participants represent different roles and are invited to contribute from various perspectives.

For example, a Circular Dialogue on a subject in a heterogeneous audience with a varying degree of familiarity on the subject could be built around the following roles (see figure 3 below):

- *Curious:* They have little experience on the subject, but want to find out more
- *Experienced:* They have some experience on the subject and are willing to share this knowledge
- *Observers:* They are asked to observe and comment, if considered useful this role can be split up further (eg the sceptics, the convinced).

The Dialogue session would take place in the following sequence:

1. It starts with a question-answer session between the Curious and the Experienced
2. Then the Observers comment on this sequence from their respective perspective
3. Next the Experienced respond to the comments of the Observers
4. At the end the Curious conclude what they have learned from the entire dialogue session. And they might be invited to start another round of dialogue by raising a new set of questions, followed by a similar sequence as above.



There are many variations to this basic sequence: Roles can be assigned according to functions which are considered useful for the given situation (as in the example above) – or they are defined in order to represent the various stakeholders. The

facilitator can either remain passive (eg time manager) or have a more active role (eg intervening to clarify questions or answers, re-focus the dialogue). As a further option, the various role holders can be given time to exchange amongst themselves and prepare their interventions collectively. Or mechanisms to document the discussion or produce a synthesis can be introduced (eg by attributing someone the role of Concluder).

But whatever the setting chosen, Circular Dialogues are a simple yet powerful tool to constructively use different perspectives. The participants are organised as a “learning system” where the resources of different viewpoints and roles are made effective. Discussing a topic with distributed roles, it is regarded consecutively from different angles, which leads to the emergence of new understanding. And to accept that different logics can co-exist and serve to interpret a given situation increases flexibility and analytic capacity.

UNDERSTANDING DIFFERENCES: REFRAMING AND CONTEXTUALISATION

Mental maps which guide the behaviour of persons or organisations also limit their choices and can block new insights and perceptions. All mental maps are representations of reality and thus necessarily differ from it. But it is important to understand which processes are involved in creating these differences (eg generalisation, distortion, deletion) and how they can be influenced so people can become open for new alternatives and solutions.

Generalisations can become very powerful filters which effectively hinder the perception of any experiences other than those expected (“self-fulfilling prophesy”). This happens when experiences become detached from the context in which they were originally experienced. But since all mental maps are only valid within a specific context, these limits need to be identified and made transparent.

Reframing circumvents the conscious control of thoughts and opens minds for new tracks towards improvement and learning. A past event which is seen as problematic is put in a new frame, so it can be viewed differently: This can be done either by placing it in a different context, where other values prevail, or by changing its meaning (eg “the good in the bad”).

Stepping out of a dominant frame of thinking is an invitation to deliberately view a situation differently and thus facilitate change. Reframing as a communication technique dates back to the original work of the Palo Alto School in the 1960s and 70s (Watzlawick et al 1974) and was later on expanded by the founders of Neuro-Linguistic Programming (NLP) (Bandler and Grinder, 1982), who focused primarily on profound analysis of language patterns and structures.

Reframing is particularly useful in situations where resistance against change is very strong or stakeholders are caught in “rigid loops”, repeated interaction patterns which they cannot change even though they might consider them dysfunctional.

The Circular Dialogues described above can also be used for reframing, especially if used in the form of role play. Participants are then invited to try out

different roles and experience for themselves (in a seemingly playful situation) how this feels – and what are the effects on others. The reframing effect is even more powerful when roles are assigned deliberately to contradict real life situations, eg if sceptics on the subject at hand are asked to act as enthusiasts.

The use of metaphors is also a very effective way to facilitate reframing. It allows people to step out of their every-day boundaries and express their experience or feelings in a different, yet common language. It is also helpful to label entire sessions in metaphoric terms (eg the kitchen, the garden, the lab), because this creates a new collective frame for joint activities. The same goes for the use of analogous (non-verbal) communication techniques, which are very helpful when dealing with relations, as they are better suited to deal with emotions and the relational aspects of communication. The most frequent ones are the use of pictures and sculptures, cartoons, jokes or sketches.

FEELING DIFFERENCES: CONSTELLATION WORK

This technique helps to reveal or transform the dynamics of social systems, which the present actors are often not aware of. Social systems are represented through the spatial distribution of people in a room, their positions and relative distance represent relations (to each other or with respect to certain issues). This way crucial aspects like proximity, distance or exclusion can be expressed non-verbally and are directly felt by the actors and their feelings are also used to identify new or more appropriate constellations. Positions can easily be changed, therefore this technique is well suited for experimenting with different options and finding new solutions.

Constellation work is based on the assumption that representatives of a social system who find themselves in a corresponding spatial configuration develop sensations and feelings which are similar to those of the individuals in the “real” situation. The perceptions of representatives are not inhibited by individual preferences or feelings and relate strictly to spatial relations and configurations. This can be particularly helpful in situations where dependencies, personal prejudice and animosities block new insights (Varga von Kibéd and Sparrer, 2000).

The example on the following page shows the transcript of a constellation session, which was used in the course of the evaluation of a business support project with the aim of improving relations among key actors: Consultants who deliver the support to target businesses, and the Agency which funds the project and provides complimentary support to businesses.

The session consisted of four sequences:

1. The initial situation: Key Consultants (C 1–3) are positioned to illustrate their current relationship (represented by distance and direction), then the two departments of the Agency (Ag 1–2), which are most relevant for the project, are positioned accordingly. Note the disconnected and sometimes even antagonistic positions of the actors.
2. The desirable future: Consultants are placed in a constellation in which they

- feel their relations are most comfortable and effective. Note their tendency to improve their connectedness and to become focused on each other.
3. The place of the Agency: The two Agency departments try out several alternative positions in relation to the Consultants – and the project, which is introduced at this stage to represent the “whole” (indicated by the arrow “project direction”). The two archetype positions are labelled “behind” and “in front” of the Consultants, and were derived from their original split positions shown in sequence 1.
 4. The solution for all: Consultants and Agency departments move around until they find a constellation which is acceptable and comfortable to all. This position is labelled “effective partnership”. Note how both Consultants and Agency departments have to modify their original desirable positions in order to find a constellation which suits all. This sequence was the most crucial, everyone involved felt immediately and very strongly how changes in the position of one actor affects all the others. And by attempting to find a solution acceptable for all they developed strong sensations for the emerging whole, represented by the project.

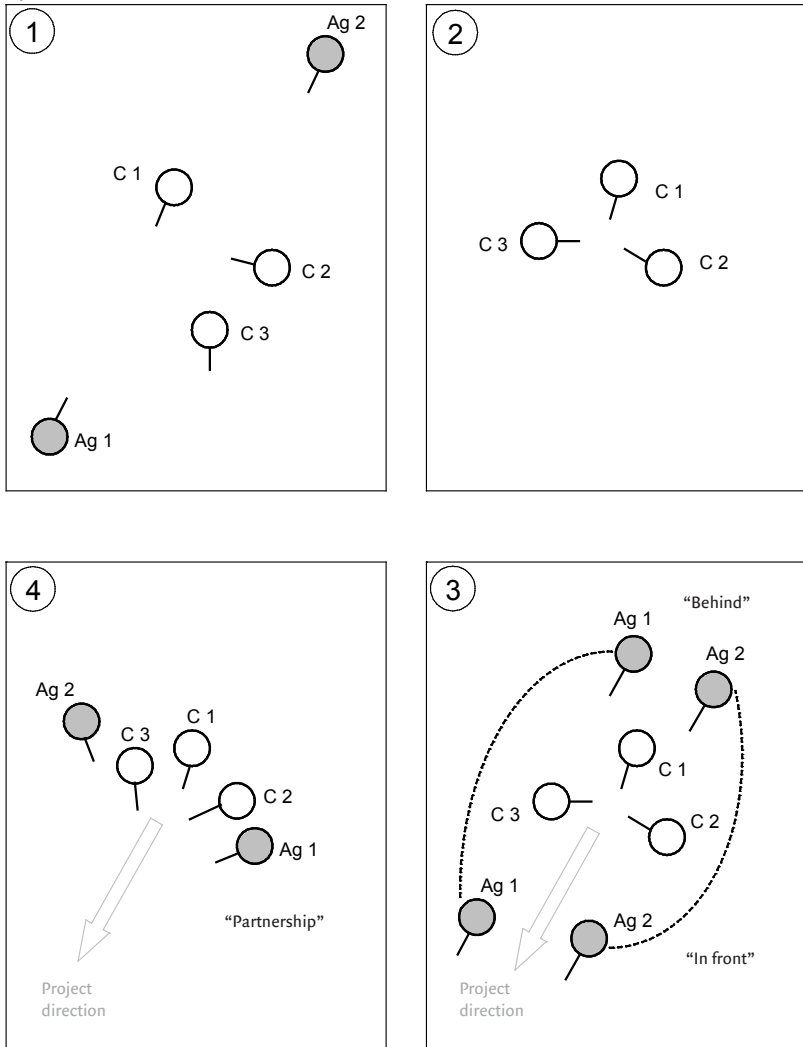
Each of these sequences is first performed non-verbally, representatives move around but once they find a firm position they are asked to express their sensations and feelings with respect to their position. This is followed by a group reflection on the actors’ constellation. The alternatives in sequence 3 are commented extensively by all, reflecting on particular strengths and weaknesses. The solution constellation found at the end is followed by a discussion which translates the findings back into the real life situation: what do these new relations mean in practical terms (behaviour, communication, activities)? how can they be achieved? what needs to be done by the actors represented? what might be the consequences for others?

Constellation work is a very rapid way (sessions last only up to two hours) to reveal hidden and largely unconscious aspects in personal relations and to test the effects of different alternatives. It allows to deal with one of the most crucial and sensitive aspects in development work – relationship of actors – in a different way, which works not primarily via language but through feelings and sensations. It makes inner maps of actors explicitly visible and brings about change by activating conscious and unconscious elements which affect people’s relationships.

This technique was originally developed for personal systems (eg families, organisations) and it has become rather popular in family therapy and organisational development. However, constellation work can also be used for larger systems, abstract systems (eg resources or development factors of a region) or for multiple representations (eg professional and personal situation of an individual or organisation).

But in any case it is advisable to use experienced constellation workers – at least in continental Europe a large number of these professionals can be found meanwhile – if one wants to go beyond gaining new insights and arrive at effective new solutions.

Figure 3



3. Utility of and requirements for systemic evaluations

The use of systemic approaches and techniques is particularly suited for the evaluation of complex realities (eg local and regional development), for evaluations which take place in cross-cultural contexts (where appropriate handling of differences in value/views is crucial) and in formative evaluations whose prime purpose is learning.

However, from a systemic perspective there are some important prerequisites for learning to take place:

- *Promoting dialogue*: First of all, learning must take place within the system that is being evaluated, facilitate the exchange of views and improve mutual understanding.
- *Integrate different perspectives*: Combine internal reflection (“self evaluations”) and external views to increase the problem-solving: Self-evaluations are well suited to reflect the complex realities and logics of involved actors. Combining them with external views can help to avoid their disadvantages and “blind spots”. External evaluators on the other hand should take care not to produce authoritative statements, but stimulate reflection and discussion.
- *Focus on utilisation*: evaluations should produce useful information, geared towards the information needs of intended users which should be identified early on.
- *Separate from auditing*: Last, but not least, evaluations which should lead to joint learning need a rather high degree of openness and trust among the involved actors. Therefore they have to be strictly separated from audit and control functions.

Another important requirement for systemic evaluations is flexibility in implementation. They are designed as iterative processes, consisting of successive reflective loops. Thus it is important for clients that they can cope with recursive designs, where only a basic outline can be defined at the start (including the available budget, a timeframe with milestones for the delivery of services or interim findings), but where the process should remain sufficiently open to respond to new findings, requirements or issues. And evaluators must maintain an overview of resources and time requirements and be able to steer the assignment within this general framework despite changing demands.

In my experience, the systemic approach is particularly relevant in situations where evaluations aim at producing learning effects beyond the level of individuals, eg at the level of projects or programmes. And which go beyond proposing short-term solutions for particular problems and aim to contribute to future successful action in complex situations. Systemic tools are important instruments to achieve such “higher level” learning processes as they allow to actually work with these differences in order to increase mutual understanding, achieve consensus or bring

forth joint solutions. They are well suited to visualize complex realities (eg causal diagrams, concept mappings, social network analysis), allow to better understand differences (eg reframing and dialogue techniques, large group interventions) or to overcome differences in innovative and often surprising manners (eg constellation work, solution focus, paradox interventions).

But applying systemic approaches also holds specific requirements for evaluators. They essentially have the role of external observers, who should not pretend to be objective, but provide additional points of view and specific skills for managing the process. But he/she is more than a facilitator, should intervene actively based on systemic principles, collect information and feed it back in varied and often surprising ways in order to trigger effective perturbations/irritations within the evaluated system, in order to find solutions or develop new patterns of interaction. An evaluator can (and should!) also express own opinions, but in an open, non-directive manner which allow for the stakeholders to chose and decide (ie present options or alternatives).

Systemic evaluators use specific skills for organising data or debates, but need not necessarily be experts in systemic tools (nor study extensive literature on their use), but they should be aware of their possibilities and know where to find suitable expertise.

Perhaps most important, they should have a general understanding of systems thinking (see section 1) and be capable to work in line with the stance outlined above:

- Conceiving the evaluand as a system is very helpful in structuring the entire evaluation, ie in reconstructing “theories of action”. This notably requires the ability to draw boundaries, identify core elements and link them in meaningful ways. It is also very helpful to involve stakeholders in these tasks and thus learn about the way they see or interpret the situation at hand.
- In order to apply systems thinking evaluators need to be good observers, capable of discovering how social systems work, ie their behaviour patterns and regulatory mechanisms. A good way to do this is by observing their reaction on past or present interventions – including one’s own during the evaluation! But this requires to be very attentive, which is particularly demanding in group settings: It is therefore advisable to work as much as possible in pairs, where one has a more active facilitation role and the other can focus on watching what happens between participants and feed-back these observations.

In my evaluation work I aim at maximum stakeholder involvement. Bringing a social system (or at least parts) together in one place allows best to trigger and observe interactions within short time spans. Thus I prefer to design evaluations as a sequence of (focus) group sessions. But in order to be effective, they must be carefully planned: who is to be involved (and who not)? which kind of assignments to give or questions to ask? in which sequence?

The systemic principles for intervention outlined above are very helpful in preparing these sessions, and in particular the formulation of some initial hypothesis about the social system(s) involved. The sessions themselves are rather short, mostly ranging from 2–4 hours. In order to trigger utmost reactions, the use of systemic tools is very effective. But once again they should be deliberately selected, based on the session's purpose and stakeholder constellation.

References

- Bandler, R and Grinder, J. 1982. *Reframing NLP and the Structure of Meaning*. Moab, Utah: Real People Press.
- Bandler, R and Grinder, J. 1975. *The Structure of Magic I*. Palo Alto, CA: Science and Behaviour Books.
- Bandler, R and Grinder, J. (1976) *The Structure of Magic II*. Palo Alto, CA: Science and Behaviour Books.
- Bohm, D. 1996. *On dialogue*. London, New York: Routledge.
- Davies, R. 1998. An evolutionary approach to facilitating organisational learning. In, *Development as Process: Concepts and Methods for Working with Complexity*. London: Routledge/ODI.
- Earl S, Carden F, and Smutylo T. 2001. *Outcome Mapping – Building Learning and Reflection into Development Programs*. Ottawa: International Development Research Centre (IDRC).
- Foerster, H von. 1970. Molecular Ethology, an Immodest Proposal for Semantic Clarification. In, Ungar G (ed). *Molecular Mechanisms and Learning*. New York: Plenum Press.
- Isaacs, W. 1999. *Dialogue and the art of thinking together: A pioneering approach to Communicating in Business and Life*. New York: Doubleday Currency.
- Königswieser, R and Exner, A. 1998. *Systemische Interventionen*. Stuttgart, Klett-Cotta
- Maturana, H and Varela, F. 1992. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Revised ed. Boston: Shambhala.
- Pawson, R and Tilley, N. 1999. *Realistic Evaluation*. London: Sage.
- Probst, G J B. 1987. *Selbst-Organisation, Ordnungsprozesse in sozialen Systemen aus ganzheitlicher Sicht*. Berlin und Hamburg: Verlag Paul Parey.
- Senge, P M, Kleiner, A, Smith, B, Roberts C, and Ross R. 1994. *The Fifth Discipline – Fieldbook*. New York: Doubleday.
- Simon, F B. 1997. *Meine Psychose, mein Fahrrad und ich: zur Selbstorganisation der Verrücktheit*. Heidelberg: Carl Auer Systeme-Verlag
- Watzlawick, P, Weakland, J, and Fisch, R. 1974. *Change – Principles of Problem Formation and Resolution*. New York: W W Norton and Co.
- Varga von Kibéd, M and Sparrer, I. 2000. *Ganz im Gegenteil – Tetralemmaarbeit und andere Grundformen Systemischer Strukturaufstellungen*. Heidelberg: Carl-Auer-Systeme-Verlag.
- Willke H. 1995. *Systemtheorie II*. Stuttgart: Gustav Fischer Verlag.

Evaluation in Complex Governance Arenas: the Potential of Large System Action Research

Danny Burns

In previous chapters you have been exposed to the smooth machine of method. Each method has been burnished over the years to give off a shiny aura of confidence and predictability. In contrast Danny's paper shows you what a systemic inquiry often feels like; a messy and sometimes confusing brew of method, inspiration, success, failure, negotiation and above all learning. Pulling people away from the traditions of linear cause and effect often creates confusion and uncertainty. Which is why, at least in Europe and Australasia, there has been a very close relationship between action research and systemic inquiry. Like many others with an action research bent Danny promotes a multi-method approach tailored to the specifics of the situation. The situations Danny describes in this chapter will be very familiar to evaluators and his reflections on those situations provide sage practical advice for evaluators wishing to become more systemic in their practice.

Glossary

CF	Communities First
NHS	National Health Service
SOLAR	Social and Organisational Learning as Action Research
WAG	Welsh Assembly Government

Introduction

This paper is grounded in the development of systemic action research within SOLAR¹ at the University of the West of England. In it I argue that the creation of dense webs of learning through a process of Large System Action Research² offers a solution to evaluators who are grappling with the difficulties of attributing causality in complex governance arenas³; the obsolescence of goals, objectives and plans in a fast moving policy environment; the myth of representativeness in diverse contexts, the limitations of “snapshots”, and the perennial problem that evaluation conclusions don't get acted upon. The paper explores the way in which emergence, resonance, and playback can underpin a flexible evaluation process

- ¹ SOLAR is a research and development team which specializes in large system action research.
- ² While there are conceptual differences between action research and action inquiry for the purposes of this paper I will use them interchangeably. I am reluctant to use the phrase “whole systems” as you can never bring everything into a learning system. In SOLAR we tend to refer either to “systemic action inquiry” or “large system action research”.
- ³ Social and Organisational systems are increasingly characterised by multi-agency interventions and partnerships. In a neighbourhood for example, a school, a Drug Action Team, a Sure Start project, a community worker, a Social Services Department, a residents' association, an advice centre etc, may all have a significant impact on any programme which is being evaluated. This makes the attribution of causality difficult if not impossible and requires us to think about different ways of evaluating.

that yields valuable systemic insights. A central concern of this work is to ensure not only that it “informs” change, but also that it creates change, and that learning is generated through that change. The paper draws on three pieces of work to illustrate the systemic concepts that underpin large-scale action inquiry work. One of these – an evaluation of the Welsh Assembly’s Communities First Programme – is explored in more detail. Here I highlight the way in which the emergent nature of both policy and practice is echoed in the emergent design of the evaluation.

Some conceptual underpinnings

SOLARs approach to social and organisational learning (Weil 2002), and to evaluation specifically, draws on thinking which has evolved from different but overlapping traditions: systemic thinking (Flood 1996, Checkland 1990, Midgley 2000); complexity theory (Stacey 2001, Shaw 2003) and action research (Reason and Bradbury 2000, Weil 1998)⁴. It is also strongly influenced by radical educationalists (Friere 1972, Boal, 1985). Like others (Checkland, 1990), I refer to *systemic thinking* rather than *systems* because the systems that we conceptualize cannot be regarded as representations of reality but rather as constructions to enable learning. Flood (1996) invites us to “study organisational forms as if they were systemic [my emphasis]”. This is not to say that the interrelationships are not real but that both their dynamics, and the boundaries, which define what is the whole that we are looking at, are open to multiple interpretations. Weil (1998) echoes this view in describing what she calls *critically reflexive action research*:

CRAR does not aim to create one representation of reality but, rather, the unravelling (and documentation) of multiple realities and rhetorics that are in mutual and simultaneous interaction (Weil, 1998, p58).

Systemic thinking then means “taking into account the whole” and seeks meaning in the complex patterning of interrelationships between people and groups of people. This highlights dynamics that are not always visible through the scrutiny of individual interactions. This is crucial because outcomes (positive or negative) will often have more to do with the interrelationship between interacting interventions than the effect of any individual action. Understanding the wider system within which these emerge is crucial because action rarely impacts in a linear way. In this context Weil (1997b) asks a crucial question for evaluators: “how do we begin to transform the perspectives of people caught up in the ‘linear archetype’ of overly simplistic cause and effect”. Systemic approaches go some way toward addressing this because they focus attention on:

- exploring the dynamics of interdependent and interacting processes.
- making more visible the effects of a whole which cannot be aggregated from the effects of the individual parts
- identifying enabling and disabling patterns and assumptions which run

4 See also other chapters in this publication.

- across the system
- bringing together unconnected contributors to connected processes in order to develop mutual understanding
- providing a framework within which evidence which is often discounted can be brought within the boundaries of the inquiry (eg emotional data, taboo subjects, power)

They also focus our attention on the small (often discounted) changes that can have a huge impact on the wider system. This is crucial for evaluation in complex governance systems.

While they no longer see the metaphor of a system as congruent with their understanding of *Complex Responsive Processes in Organisations* the work of Ralph Stacey and Patricia Shaw has been immensely influential in the development of our thinking. They offer crucial insights into where learning happens, how knowledge travels, and how innovation is created and embedded. This *complexity*-based approach collapses the distinction between action and decision-making, as emergent understandings fashion new pathways for action in the “real time” of their creation. This has important implications for evaluation and requires us to make a crucial distinction between formative evaluation understood as a process of real time feedback to decision makers who can set change in motion, and a more organic process through which a myriad of different stimuli (including those introduced from the top – policy, audit etc) lead to changing relationships and new sense making across a diverse terrain. Sense making comes about through engaging with others in conversation. Action that is embodied in changed practices and norms occurs as a result of these multiple conversations, and becomes embedded through emergent self-organisation. Shaw describes this as an improvisational process:

“the living craft of participating as an intentional fellow sense-maker in conversation after conversation (both public and imagined), encounter after encounter, activity after activity. I want to help us appreciate ourselves as fellow improvisers in ensemble work, constantly constructing the future and our part in it as daily activity as we convene or join or unexpectedly find ourselves in conversations” (Shaw 2003, p172).

Action Research is also based on the interplay of dialogue and action and emphasizes the importance of *emergence* in sense making. But to support change across an organisational system requires us to build considerably on dominant action research models. Action Research often involves a group of “stakeholders” engaged in an inquiry together, over a perhaps a year, about an issue⁵. They are participants in continuous cycles of analysis, reflection, planning and action. This

5 A good general introduction to action research is Greenwood, D and Levin, M 1998. Those interested in a more extensive exploration of the theory, practise and history of action research should start with Reason and Bradbury’s (2000) *Handbook of Action Research*.

practice is now well established and in my view has robustly answered the many methodological questions which have been asked of it with regard to validity, quality, researcher neutrality, generalisability and so on (Lincoln 1995; Bradbury and Reason, 2000).

What is more problematic is that action research – particularly cooperative inquiry (Heron 1996, Reason and Rowan 1991) – is often characterised by small groups, with inquiry centred on a single group (eg a management team, a community group, an action learning or cooperation inquiry group etc), and the boundaries of the group are often relatively closed. Furthermore, implicit in the inquiry process is the assumption that the group or individuals within it can enact changes that result from their deliberations. But our observation of both action research and action learning is that frequently resistances occur because the group and the emergent sense making are separated out from practise.

One alternative to small group action research has been the development of large events and conferences that bring the “whole system” into the room. Open space (Owen 1992), Future Search (Weisbord and Janoff 1992), and World Café (Brown 2001) are all approaches which we have drawn on. These have a strong orientation toward action but because they are usually one off events they too tend to individualize action, and responsibility for action tends to be vested in the commissioners or facilitators of the events. I have often experienced a great sense of breakthrough at these events, but after a few weeks the details recede, and it is difficult to convey exactly what happened and its significance to others who were not there. Some of the most profound conversations have been the most difficult to convey. Shaw contrasts her approach to this:

The open space event generates a strong temporary sense of community, whereas the kind of work I am describing generates a rather weaker, shifting, ill – defined sense of “us” because conversations are always following on from previous conversations and moving on into further conversations involving others. People are often gathering and conversing around ill-defined issues, legitimation is often ambiguous, motivation is very varied. The work has much less clear and well-managed beginnings and endings. There is not the same sense of creating common ground for concerted action. There is no preconceived design for the pattern of work; it evolves live. We are not necessarily trying to create outputs in the form of public action plans; rather, we are making further sense of complex situations always open to further sense – making and in so doing redirecting our energies and actions (Shaw 2002, p146).

Our approach has been to catalyse conversations by growing multiple inquiry groups interspersed with large group processes across the breadth and depth of an organisational system. This can involve between five and twenty (or even more) inquiry processes evolving in parallel, and ensures that the large event

becomes part of an ongoing process of inquiry⁶. In this way systemic mapping, reflection and action have a temporal as well as a spatial dimension. The issues raised by the groups are not pre-determined, nor are they pushed too fast toward outcomes, but there is a more deliberate process of steering by facilitators who bring different constellations of people together, connect issues that are arising in different parts of the system, help to identify key questions, support the mapping and tracking of both emergent insight and action, and so on. This places greater emphasis on the role of facilitators in “holding” a complex multi stranded process. Deliberate facilitation also enables us to interrupt flows of power that are embedded in the discourses of the “status quo”. Without this, “self-organisation” is steered by these discourses. A good illustration of this is the way in which the physical regeneration discourse continuously reasserts itself within community regeneration programmes.

A facilitator who is supporting systemic action inquiry can also help to set up inquiries/ conversations/ dialogues that might not otherwise happen. In some of our earlier projects we did not realise the potential for this. Some years ago we were asked by an inner London Community Health Council to facilitate an inquiry process with young people about health in their community (Percy–Smith et al, 2003).

Following six months of peer research where young people engaged in dialogue with other young people about health, we organised a large event bringing around 80 of them together with around 30 professionals, managers and policy makers. I recall a half hour discussion with six young people on sex education. We talked about their experiences of sex, how they learned about it, how it had been dealt with at school, and why sex education was so bad. Later in the day we asked senior managers to reflect on what they had learned. The Director of Education stood up and gave an honest account of the difficulties they were having in response to drugs problems and then, almost as an aside, said that perhaps they needed to learn from their successful sex education program. I immediately picked up the roving microphone and took it to table where the young people who I had been talking with earlier were sitting. They talked to the whole conference about the disjuncture between their experience and that of the professionals.

I now think that we should have secured agreement then and there for the Director of Education and other professional colleagues to continue to work with these young people, in an ongoing inquiry group, to co-construct and enact a new approach to sex education.⁷ In our Bristol Children’s Initiatives project we designed a process with a number of parallel Inquiry streams. In one inquiry

6 Gustavsen has recently articulated the need to extend the action research process so that “Instead of using much resources in a single spot to pursue things into a continuously higher degree of detail in this spot, resources are spread over a much larger terrain to intervene in as many places in the overall movement as possible.” (Gustavsen, 2003: 96–7). Yoland Wadsworth’s work on enhancing mental health services (Wadsworth 2001) which incorporated 35 smaller action research projects is one of the few examples of this approach enacted.

7 This also reflects the “Open Space” principle that the right people are the people who are there.

on domestic violence a conversation emerged about the role of men as fathers, these issues also emerged in the inquiry streams on “community participation”, and “childcare”. I now think we needed to actively stimulate a new inquiry on “fathering”, and because we didn’t, the issue was only ever dealt with tangentially. The active nurturing and development of emergent inquiry is a fundamental part of the large system process.

How systemic thinking can enhance evaluation

Before exploring the shape of a large system inquiry in detail I want to illustrate the importance of systemic thinking with reference to two smaller examples.

In 2001, I was working with a learning group that was made up of the Senior Management Teams of a City Council Education and Social Services Departments and the two city Primary Care Trusts⁸. They got to talking about care pathways for older people, and were musing on a puzzle that they had observed. This was that there seemed to be a one directional path from caring for yourself in the home, to homecare, to residential care, to nursing care, and ultimately to acute hospital beds. If someone had to go to hospital (for example after a fall), they almost invariably ended up staying there longer than anticipated, and from there, instead of going back home, or back to residential care, they ended up in nursing care. Some of them started to talk about the experiences of their own parents. One talked of the way in which older people who came into the wards were routinely catheterised. This was because it took 15 minutes to take each patient to the toilet and 15 minutes to take them back again. There simply wasn’t enough nurse time to do this. People in the room realized that this was not a unique experience and its effect was that virtually all of the older people who came onto the wards as independent came out as dependent. So they could no longer go back into their homes or into residential care. This meant that there was a much greater demand for nursing care within the social care system as a whole (which could not be met). This in turn meant that people had to stay in hospital for longer because there was nowhere for them to go, thereby exacerbating the “bed blocking” crisis within the National Health Service (Burns 2003).

A report on hospital bed management showed that two million bed days had been lost each year because of delays in discharging people who were fit to leave hospital. People over the age of 65 occupied two thirds of the beds, and a key factor in their delayed discharge was the difficulty in finding them places in community facilities. *The Guardian* 17th April 2002

This story not only illustrates for me the importance of systemic work across organisational boundaries, but also the centrality to evaluation of what Midgley (2000) calls boundary critique. In other words what you include within your system

8 Primary Care Trusts are the part of the UK NHS that manages primary care services (GP services, etc and a range of other community services such as public health).

boundary fundamentally affects the outcome of any assessment/judgement. Where do we place the boundary in this example? If we assess the micro ward practises solely in the context of a budgetary constrained clinical environment, we may conclude that the ward is efficiently managed. If we draw into our analysis the consequences of dependence for the patient we might conclude that the cost of that efficiency is too high. If the boundary is taken wider to include the social care system, then not only does it become apparent that there is a negative impact on the social care system, but this is leading to a much higher cost for the hospital. If you asked the nurse on the ward about these effects s/he wouldn't be able to see them. If you asked health service managers what was causing the bed-blocking problem it is unlikely that s/he would be able to locate its roots in catheterisation. If you were a social care manager it is unlikely that you would know anything of what was happening on the wards unless you had been personally affected in some way. It seems to me that a crucial role for an evaluator is to open up these questions.

My second example comes from an evaluation of a programme set up to fund and support community based children's projects in "deprived" city neighbourhoods. SOLAR did some work with the management board on how they understood the programme that they were managing. We asked them to draw pictures of the system. The pictures included:

- Divers, with diving suits and breathing equipment. Others on the surface refusing to immerse themselves.
- A beekeeper dressed in protective clothing, collecting honey from the beehives surrounded by hundreds off swarming bees.
- Three planners looking at three different images of the building site that is in front of them. The site is being knocked down as it is being reconstructed.

Through discussion new interpretations of the system dynamics emerged. In many of the pictures the environment, which holds the children, is depicted as dangerous. The "workers" are all wearing protective suits to engage with the children. What does this tell us about the assumptions we hold about the children that we are working for? What does it tell us about the relationship between senior managers and the children that they are there to serve? These questions and analysis invite us to ask whether a particular approach to service delivery is being adopted because of the conception of the systemic whole that is being held. The drawing with the bees also illustrated a uni-directional process whereby the product of the bees was reaped by the workers, who passed it to people further up the system, who in turn passed it up to a figure with a crown on his head. This chimed in with another drawing that had depicted the organisation as a bank.

Other questions start to arise: Was it possible that the system was in practice oriented to make a profit on its investment rather than meet it's espoused purpose of supporting the inclusion of "disaffected" and vulnerable young people (Argyris and Schon 1974). How we conceptualise a social and organisational system has an

impact on how we structure it and how we intervene in it, and again this is crucial material for evaluators to engage with. It is important to note that these examples illustrate how systemic thinking can be applied to a variety of contexts and scales – not only to large programme evaluations.

A large scale emergent evaluation

So far I have illustrated some different ways in which strategic insight can be generated with a small group. In this section I explore a large system inquiry for the evaluation of the Welsh Assembly's communities First Programme. Communities First is described by the Welsh Assembly as its flagship programme. This has meant that it is very much in the spotlight and is operating in a highly politicised context. It was conceived as a ten to fifteen year programme – with £83 million set aside to support 142 neighbourhoods in the first three years. Each project has a coordinator, and most commonly their first task has been to build a formal partnership of stakeholders to support their work.

The projects are seen as catalysts for community based regeneration, and their work centres around capacity building, networking with a view to influencing mainstream programmes, local needs identification, leveraging larger scale regeneration monies into neighbourhoods, and managing some capacity building services. Their power lies mainly in their capacity to influence and in growing new forms of distributed leadership that can exert their own pressure for change. Situated within the Communities Directorate of the Welsh Assembly, the programme is administered by a core team in Cardiff and five regional implementation teams. A £2 million annual support package was put in place. Unlike most other UK regeneration programmes it has been highly non-prescriptive which has given it great potential for innovation. It has also been underpinned by strong Community Development values. The programme has been running for four years and SOLAR has been working on it for more than two. This paper describes an ongoing process.

We initially designed the evaluation with three parallel strands. An action research element which would involve three detailed action research projects, case studies of 30 partnerships, and household surveys which would be carried out in all 142 areas by a social survey company. SOLAR was responsible for the action research part of the programme.

Our proposal for the action research was firstly to carry out exploratory interviews in each area, secondly to support a peer research process built on themes that had emerged from the interviews, thirdly to carry out a sequence of multi stakeholder conferences structured around the peer research outcomes, and fourthly to set up inquiry strands to pick up key questions which needed resolution. We felt that the depth of insight gained through the depth inquiry processes in three areas would complement the snapshot data from the case studies and the survey based data across the whole programme to give a more complete picture of the whole. What it didn't do was offer us enough scope to

map the patterns and inter-linkages, which were emerging across the whole programme, nor did it connect us to the diverse sources of power, which could create change within the system.

However, following early conversations with officials within the Welsh Assembly Government, we were able to add a systemic action research group at the centre with all of the main players involved in the management and governance of the programme: the head of the Communities Directorate: the head of the Communities First programme: all of the officers based within the Welsh Assembly: the leads of the five regional implementation teams, and two representatives from the Communities First support Network. The group now meets for a whole day every six to eight weeks and considers this event to be a crucial part of their policy development and implementation process. The creation of the WAG Action Research Group marked a shift in the focus of the evaluation so that instead of three parallel strands (a) case studies (b) survey data (c) action research, the action research process became a hub through which learning about the programme was analysed and acted upon.

By now it was becoming clear that we also needed some sort of inquiry process that could form a bridge between the Welsh Assembly at the centre and the 142 projects. So we set up three *cluster groups* (of rural partnerships, welsh valley partnerships, and urban partnerships). These took the form of facilitated one day events which were a forum (a) for testing the resonance of issues which had emerged elsewhere the learning system and (b) for collective dialogue *ffi* making visible new issues and generating new evidence.

One issue that emerged concerned “the relationship between CF partnerships, and local councils”. Although the Welsh Assembly funded the programme the money was being channelled through local “grant recipient bodies” (mostly local councils) who also acted as employers etc. Cluster group meetings highlighted difficulties in the relationship between some local councils and CF partnerships. It was clear that action needed to be taken to strengthen relationships, but it was not clear what they should be. So we moved to set up a *strategic action inquiry group* involving all of the key stakeholders on this issue. The emergence of this process reflects a widening of the system boundary with the realisation that learning confined to the WAG/ CF project axis would not be enough.

Meanwhile, the *local action Inquiry groups*, which we had set up on the ground, expanded in number and changed their focus to become more thematic. They remained rooted in the experience of specific localities, but extended their scope to explore issues (youth, embedding local evaluation, relationships with local councils, equality, and “mainstream programme bending” etc) across the programme as a whole. The most recent evolution of the process has been to develop an extensive mind map containing “data” generated by the evaluation (including “stakeholder” and “evaluator” perceptions and interpretations) allowing participants to interrogate, learn from and challenge emergent “findings”.

The outcome of all of this is a dense network of action inquiries. Some of these have started in an open ended way, others are charged with problem solving, Most identify new issues which can in turn open up new inquiries and catalyse new leadership. As the network of people involved in the evaluation process becomes wider, the process becomes more participative and more accountable. Some examples of the changes that have resulted from the inquiry process are:

- the restructuring and re-profiling of a £2 million per year support programme
- priority attention shifting to relationships with local authorities
- clarity about the capacity building focus of the programme leading to clear funding guidelines
- the co-creation of intermediate outcomes by which to judge the impact of the programme after five years
- re-writing of the programme guidance.

Alongside all of this we continue to carry out traditional case studies. The team is carrying out a follow up survey of co-ordinators who have left, and a baseline assessment of local “social capital” in 10 neighbourhoods. The survey resource has been redirected to focus on a detailed understanding of these neighbourhoods. All of this feeds into and supports the action inquiry process.

Designing large system evaluation

Having described the shape of the process, it may be helpful to note a number of conclusions that we have drawn from our practice.

EMERGENT EVALUATION

Emergent policy and practice require an emergent evaluation. This can be simply illustrated by a Communities First example. We discovered fairly early on that many stakeholders conceptualised Communities First as a mainstream regeneration programme rather than a capacity building programme⁹. This led to some confusion about direction and focus in the early stages (National Assembly for Wales, 2005). We brought this tension to the “WAG Action Research” group who spent some time clarifying the core focus, and translating it into explicit funding guidelines.

The process involved much more than “goal clarification”. A systemic perspective was necessary to work through the patterns that had contributed to the lack of clarity. Embedded norms (resulting, for example, from professionals who had previously been employment in traditional regeneration work) and competing drivers for change were amongst the many factors impacting on the situation. For instance, some politicians and residents wanted to see “quick wins”; some local

9 £83 million spread across 142 areas over 3 years averages out at roughly £200,000 per year. This buys a lot of community development and ‘capacity building’ but very little physical infrastructure.

authorities saw CF as little more than an additional pot of money to add to their existing regeneration programmes, and so on.

A similar process of exploration needed to be undertaken in relationship to “mainstream programme bending”. While the broad aim that CF should impact on the way in which mainstream services were delivered was always present, until the programme was up and running it was impossible to know what this might look like and what was possible/realistic. It was only having gone through a process of inquiry around these issues that that we knew exactly what outcomes we were trying to evaluate. And as these things become clearer so the design of the evaluation evolved, resulting in for example, the creation of participatory processes to establish intermediate outcomes upon which the programme could be assessed after five years. As the programme defined itself in practice the evaluation had to be flexible enough to respond.

The process that I described above evolved over nearly two years. We could not have sustained the core action research group unless we had built the trust of the senior WAG officials. Once we knew a large system approach had been embraced it was possible to consider extending the focus of the action research groups. The groups (and the case studies) were generating lots of issues that needed to be tested across the system so we set up the clusters groups. The cluster groups (amongst other things) highlighted the need for coordinators to be engaged with the evaluation material that was emerging so the mapping tool emerged as a solution. One thing led to another. So just as the evaluation design is highly emergent so are the research methods that have been selected in response to the emergent evaluation design.

METHODOLOGY AND METHOD

It is important to distinguish between methodology and method. The methodology I have described is what we might call “systemic action inquiry”. The “methods” are varied and may include some quite traditional qualitative research activities such as semi structured interviews and participant observation, as well as those more closely associated with action research (inquiry groups, peer research processes, large events, and interactive mapping). Rather than trying to replicate method across the system to ensure reliability, generalisability, comparability etc. we deliberately introduce mixed methods in order to cast different lights on the dynamics of emerging problems. The use of multiple methods enhances the scope for insight generation and enables important data, which is often discounted in traditional evaluations to be surfaced. The challenge is to juxtapose these methods in ways that shed more light on the dynamics of interventions.¹⁰

Rigorous evidence based on action inquiry requires a dense network of inquiry groups through which insight can be generated and corroborated. The quality of the meaning making process (analysis) lies in the exposure of information, reflections and interpretations to ever increasing circles of peer review at multiple

10 See Gerald Midgley's and Ken Meter's chapters for further discussion of mixed methods.

levels within the system. What should be clear by now is that the methodology is the gradual creation of a dense action inquiry system. In contrast to the traditional process of collecting and analysing data and then disseminating it, the inquiry and resultant action is part of the data, and its journey through the networks of individuals and groups is the dissemination.

In this approach there is no formal system modelling or process-modelling characteristic of much systems work. This is an important point because although large system inquiry facilitators will want to take into account many of the key factors identified in for example Soft Systems Methodology or Total Systems Intervention, their systematic approach runs counter to the ethos of our work. We have found that many of the people we have worked with have been happy to discuss issues, tell their stories, and even try to understand the systemic dynamics of which they are a part, but begin to withdraw as soon as the process becomes too “modelled”. I suspect this is partly because the model can never do justice to the complexity of their reality. It is also because working through the detailed steps of a complex model will often fail to sustain the interest of non-researchers. Maintaining engagement is crucial, because it is participation that provides the underlying legitimacy and validity for the analysis. The sense making process needs to be as “accessible” as the data collection process, enabling stakeholders to carry out their analysis in ways that hold meaning for them. As a result each group may approach its issues and problems in a different way. The methods used may be driven by the participative process, by the intuition of the facilitators, through opportunistic connections to things that were happening anyway. Equally, they may follow the logic of emerging insight (eg now we have got to this point, we need to ... “interview x people” or “see what happens if...”). Interpretations are strengthened as they travel through the different groups. Their strength can at least in part be judged by the success or otherwise of the actions that emanate from them.

FOCUS ON RESONANCE NOT REPRESENTATIVENESS

Resonance is a crucial organising principle for systemic work in general, and is likely to be more reliable than representativeness in large system contexts. Wherever we are working with a group on issues that people think may be a fractal of the wider systemic pattern they can firstly be played back into their source environment. Then we can play them into cluster meetings. If there is a strong resonance at cluster level, then we can take it out to whole system level. This enables us to deal with the problem of scale. By testing the resonance of an issue we are testing the strength of feeling and importance of an issue rather than its frequency (which may or may not have significance). Understanding significance involves witnessing the emotion that is attached to issues, watching what people are drawn towards, what they will fight to express and what they will let go. This can only be done in-group dialogue. This gives robustness to the systemic analysis, and in conventional terms can be seen as a form of triangulation because issues identified in one part of the learning system are checked for relevance in the wider system.

BLURRING THE DISTINCTION BETWEEN EVALUATION AND DISSEMINATION

Traditional evaluation distinguishes sharply between the evaluation, the product of the evaluation and the dissemination of that product. Large system action research sees them as one and the same thing because the networks of inquiry are the paths by which dissemination is travelling across the system, and the “findings” of the inquiries and the consequent actions taken are also the subject of ongoing inquiry. This of course has significant implications for reporting. In a large system context it is crucial because experience tells us (and this has been corroborated by interviews in the CF context) that written guidance, best practice and case study materials do not easily translate into learning.

Conclusions

Work of this sort requires research commissioners and clients to develop a different outlook on what evaluation is, and to challenge considerably their assumptions about it, and expectations of it. A large system action research evaluation collapses the distinction between evaluation, policy and practice development.

Dissemination becomes part of the process of data collection and analysis. At the contracting stage the client has to commit to a process whose outcomes and even whose form will be at least partially unknown. Many clients still look for the reassurance of a “conclusive” as opposed to “continuous” product but in this work formal “reporting” is likely to be a less significant.

There is a level of risk in the work. Some inquiries will produce incredibly rich data and transforming action as a result, but not all will, and sometimes the inquiries won’t always relate to what commissioners think they want evaluated! Some inquiries stall, and this can be complex in a world where every element of public programmes is individually performance managed. I hope however that I have illustrated the huge potential of a changed outlook on evaluation, and a realistic process for delivering it. When so much contemporary evaluation work is focused on outcomes it is ironic that, we often struggle to demonstrate the outcomes of our evaluations. This is one of the greatest strengths of systemic action research. We can show through the long list of actions that emerge from the work exactly what the outcomes of our evaluation are. Far better than an authoritative report that sits on the shelf!

References

- Argyris, C and Schon, D. 1974. *Theory in Practice: Increasing Professional Effectiveness*. San Francisco: Jossey Bass.
- Boal, A. 1985. *Theatre of the oppressed*. Theatre Communications Group: New York
- Burns, D. 2003. *Whole System Action Research in Complex Governance Settings*. Key note presentation to the ALARPM world congress.

- Bradbury, M and Reason, P. 2000. Broadening the bandwidth of validity: Issues and choice points for improving the quality of Action Research. In, Reason, P and Bradbury, H, 2000, *Handbook of Action Research*. London: Sage.
- Brown, J. 2001. *The World Café: Living Knowledge Through Conversations that Matter*. Doctoral dissertation available through Pegasus Communications.
- Checkland, P and Scholes, J. 1990, reprinted with a thirty-year retrospective 2004. *Soft Systems Methodology in Action*. Chichester: John Wiley.
- Flood, R L. 1996. Total Systems Intervention: Local Systemic Intervention. In, Romm, N and Flood, R. 1996. *Critical Systems Thinking: Current Research and Practise*. New York: Plenum.
- Greenwood, D and Levin, M. 1998. *An Introduction to Action Research: Social Research for Social Change*. Thousand Oaks CA : Sage Publications.
- Gustavsen, B. 2003. Action Research and the Problem of the Single Case. *Concepts and Transformation*, 8(1): 93–99.
- Heron, J. 1996. *Co-operative Inquiry: Research into the human condition*. London: Sage.
- Lincoln, Y. 1995. 'Emerging Criteria for Quality in Qualitative and Interpretive Research. In, *Qualitative Inquiry*, 1(3) 1995: 275–289
- Midgley, G. 2000. *Systemic Intervention: Philosophy, Methodology and Practice*. New York: Kluwer academic.
- National Assembly for Wales 2005. *Minutes of the Social Justice and Regeneration Committee 6 July 2005*
- Owen, H. 1992, 2nd edition 1997. *Open Space Technology: A Users Guide*. San Francisco CA: Berrett-Koehler.
- Percy Smith, P, Burns, D, Walsh, D and Weil, S. 2003. *Mind the Gap: Healthy Futures for Young People in Hounslow*. Bristol, UWE and London: Hounslow Community Health Council.
- Reason, P and Bradbury, H. 1990. *Handbook of Action Research*, London: Sage
- Reason, P and Rowan, J. 1981. *Human Inquiry: A Sourcebook of New Paradigm Research*. Chichester: Wiley.
- Romm, N and Flood, R. 1996. *Critical Systems Thinking: Current Research and Practice*. New York: Plenum.
- Shaw, P. 2002. *Changing Conversations in Organisations*. London: Routledge.
- Stacey, R. 2001. *Complex Responsive Processes in Organizations: Learning and Knowledge Creation*. London: Routledge.
- Weil, S. 1997b. Social and Organisational Learning and Unlearning in a Different Key: An introduction to the principles of Critical Learning Theatre and Dialectical Inquiry. In, Stowell, F et al (eds) 1997. *Systems for Sustainability: People, Organisations and Environments*. New York: Plenum.
- Weil, S. 2002. Learning from the Evidence of Co-inquiry Based Practice and Research: Explorations in Primary Care. In Brockbank, A, et al 2002. *Reflective Learning in Practice*, Aldershot, UK: Gower.
- Weil, S. 1998. Rhetorics and realities in public service organisations: systemic practice and organisational learning as Critically Reflexive Action Research (CRAR): *Systemic Practice and Action Research*, 11: 37–45
- Wadsworth, Y (ed). 2001. *The Essential U&I*. Carlton South, Victoria: VicHeath.
- Weisbord M R and Janoff, S. 1995, 2nd edition 2000. *Future Search: An action Guide to Finding Common Ground in Organisations*. San Francisco CA: Berrett-Koehler.

Acknowledgements

I would like to offer special thanks to my colleague Susan Weil with whom many of the ideas that have underpinned this work have been co-developed; I would also like to acknowledge the important contributions of Dianne Walsh and Barry Percy-Smith who have been working with me as action researchers on the Communities First programme; and Matthieu Daum who was an action researcher on the Bristol Children's Initiative project. Thanks also to the many people who commented on this paper: Bob Williams, Bob Dick, Gerald Midgley, Meenakshi Sankar, and Iraj Imam who in reviewing the first draft provided invaluable incisive comment and Yoland Wadsworth whose reflections were immensely encouraging.

USING SYSTEMS CONCEPTS IN EVALUATION

Evolutionary and Behavioral Characteristics of Systems

Jay Forrest

Like others in this volume, Jay's chapter reveals more every time you read it. What he lays before us is a true heuristic – a system of systems concepts that as you step through it reveals more about how to conduct yourself and how to conduct your evaluation. As evaluators, we go out, observe situations and make judgements about those situations based on sets of assumptions. Jay's heuristic provides us with ways by which we can reveal those assumptions and assess the extent to which they may or may not be correct. Furthermore he gives us the means by which we can replace those assumptions with a more accurate or appropriate set. If this sounds easy, it isn't – but a careful reading of Jay's ideas can certainly make it clearer and more robust.

Overview

The evaluation process of assessing the strengths and weaknesses of programs, policy, personnel, products and organizations to improve effectiveness not only suggests consideration of the context within which the subject of evaluation operates, but also implies some level of validity into the future. As a practitioner of both system dynamics and foresight I have long struggled to maximize insights both the context within which issues reside and the alternative futures that feasibly evolve from the present – to bound the future in a beneficial manner. As a doctoral student, I developed the methodology for qualitative system analysis that is presented in this chapter. This methodology uses qualitative system characteristics to infer likely patterns of future system behavior and evolution. The process stimulates examination of perceptions and their implications in a way that strives to lead users to a stronger understanding of not only the future uncertainties surrounding the topic of study, but also peripheral issues which may impact on the topic.

This methodology begins with a qualitative, systemic view of an issue or topic for evaluation and focuses on perceived characteristics of the system. The process stimulates closure between perceptions and their implications, builds a stronger understanding of the topic under study and of the alternatives for proactive action.

As this approach deals with perceived characteristics the process benefits from surfacing and considering diverse perspectives and inputs during the development of the system description – typically in a facilitated process of developing a qualitative system diagram (causal map, causal loop diagram, or influence diagram) of the item or issue under consideration. The results are not as determinate as for quantitative modeling but the process has three key benefits:

- A deeper understanding of the system and its environment
- A logical basis for establishing the boundaries of a model, and
- A logical basis for inferring likely patterns of future structural change.

This methodology is particularly helpful for highly complex problems and issues where the detailed nature of causal links is likely to be controversial and where longer-term insights or an extended model life are desired. This chapter provides an introduction to the process and underlying logic.

The approach begins by considering the system from a perspective of maturity and proceeds by viewing the problem or issue from a series of qualitative perspectives that mutually inform the process to stimulate deeper insights into the system. This process is cyclic and repetitive – insights gained from each perspective provide input to and test the interpretation of other perspectives. The resulting insights raise new questions and stimulate revisiting prior perspectives to resolve conflicts and to elaborate the mental model of the system.

This methodology is particularly appropriate at the beginning of an evaluation as it focuses on the issue as a whole, builds understanding of the issue and its environment, and provides insights into the implications of perceptions regarding the future of the issue. The resulting insights augment other qualitative and quantitative systems approaches. An enhanced understanding of boundary issues and uncertainties provide a basis for development of stronger, more robust mental models and results. This methodology is particularly appropriate for groups, where it stimulates the surfacing of differences of perception and the building of a shared group model. While this approach uses language and insights from a variety of disciplines the concepts are presented using the language of system dynamics and the concepts of stocks and flows.¹

Foundations of This Approach

The eight perspectives used in this methodology are derived from research and fundamental concepts from a variety of disciplines. I assimilated those perspectives and relationships into the presented logic. The six key concepts underlying this methodology and their foundation disciplines follow:

1. Stocks drive systems (System Dynamics).
2. Feedback Loops serve as long-term (or primary) drivers of systems and provide leverage for influencing the behavior of a system (System Dynamics, Electrical Engineering).
3. System structure influences system behavior (System Dynamics, the work of Michel Godet).
4. Patterns of system structure (ie connectivity patterns) and evolutionary tendencies are a function of the maturity of the system (Biology, Ecology, and Biochemistry).
5. The dynamic equilibrium of an evolutionary system is a function of both the maturity of the system and the stability of the system and its environment (Evolutionary Ecology).

¹ Many of the examples used in this chapter are business oriented. If the concept of suppliers doesn't work for you, try substituting sources of communication, sources of ideas, or sources of money.

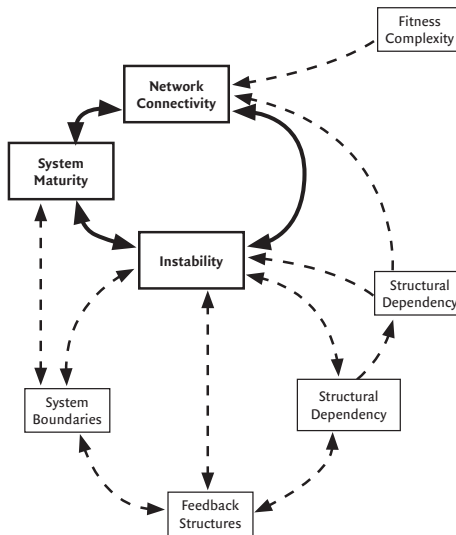
6. Fitness in a fitness landscape and the resultant pattern of possible evolution is a function of the level of complexity of the landscape. This is further influenced by the complexity of the number of factors determining fitness (“fitness function”), or in other words, the interconnectedness of fitness (Mathematical Biology and Self Organization).

You may find some of this material challenging due to the range of disciplines underlying the logic – particularly given the limitations of this volume. However, fluency in the ideas is not necessary to apply these concepts productively. Several of the concepts typically take some time to come into focus for new users. Additional material is available at my website <http://www.jayforrest.com> for those who want to explore in more detail.

The Methodology

The heart of the methodology lies in a dynamic balance between the System Maturity, Instability, and Network Connectivity patterns as emphasized in bold in Figure 1. Five other perspectives provide insights for framing, testing, and informing creating the cyclic process. The five supporting perspectives and their relationships are shown in lighter print and dashed lines to emphasize their qualifying role to the key dynamic balance of maturity, instability, and connectivity. The framework is provided to aid in connecting the concepts as they are introduced in the following sections.

Figure 1: Eight Perspectives for Qualitative System Analysis



1. SYSTEM MATURITY

One begins this process by assessing the maturity of the system under study. Real-world systems – whether economic, ecological, organizational, or social – routinely follow some variation of an S-shaped development path similar to that in Figure 2. Understanding the four phases of maturation and the associated patterns of structural change provides insight into likely future patterns of change.

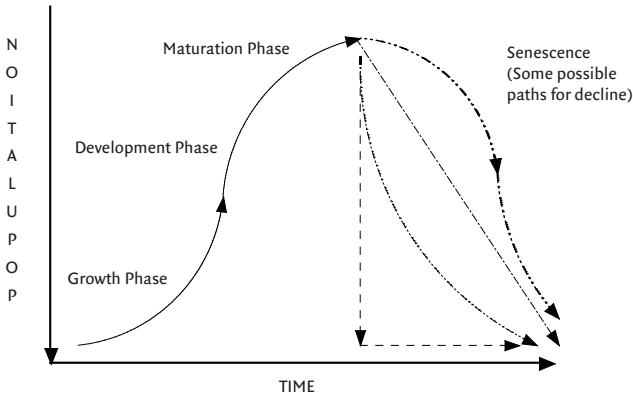


Figure 2: A Typical S-Shaped Growth Curve (Forrest, 2004)

Contemplating the system from the perspective of the classic S-shaped growth curve is often a good starting point for this process. How old is the system? Does the system, organization, program, funder, or issue appear to be on the immature, developing side of the curve or does it seem to be peaking, with flattening growth, somewhere in between, or declining? Immature systems are often relatively young, experimental, rapidly growing, and dynamic. More mature systems are generally older, with more experience, more optimized, more standardized and less experimental, and often more bureaucratic. Assessing the state of a system is often not clear-cut for different parts of the aggregate system may be in different stages and instabilities complicate the assessment. It is easiest to grasp the systemic behavior of systems in a stable environment. The following section describes the typical maturation phases and process to serve as a basis for elaboration by subsequent perspectives. The names and characteristics of the maturation phases follow the generic scenario of growth, development, maturation, and senescence for living systems described by Ulanowicz (Ulanowicz, 1997).

The Growth Phase

The growth phase typically begins immediately after a system² undergoes a major destructive perturbation or when the system enters a new domain. Events

² Businesses, nonprofit organizations, causes, families, or even lives can be viewed from a systems perspective.

initiating growth phases commonly include catastrophes, inventions (including new concepts and programs), technological breakthroughs, physical relocation or expansion, and elections. In the growth phase there is generally little direct competition and necessary resources are generally adequate or readily available. During the growth phase stocks typically display rapid or exponential growth (within the bounds of enabling resources). The growth phase extends from initiation to the area of the inflection point on a typical S-curve growth pattern. During the growth phase actors in the system generally focus more on growth and opportunity (exploiting available resources) than upon efficiency. During this phase the actors in the system are often striving to find “better answers.” New connections (formulas, relationships, configurations, etc.) are formed and tested. Older connections (products, concepts, relationships) are pared as better combinations are found. Systems often fail during the growth phase if successful combinations are not found. The system structure is likely to be relatively turbulent early in the growth phase as new, “more successful” combinations are identified. Over time the emphasis generally shifts to repeating or replicating the successful formula and the system enters the development phase. From a systemic viewpoint the growth phase begins as the search for successful structure and relationships and continues through the replication or exploiting of that successful structure. Successful systems eventually encounter resource limitations and/or competition for resources leading to the development phase. Examples of system turbulence during the growth phase might be related to competition (both idealogic and/or institutional), challenges of growth, resistance to change, and shortages of talent, clients, or political support.

The Development Phase

The transition into the development phase is marked by a shift from emphasis on finding the best answers to replicating successful formulas and building infrastructure to exploit opportunities. During this phase growth begins to slow – typically as a result of declining resource availability, program beneficiaries, or growing competition from other projects or programs. Declining resources encourage efficiency and actors within the system typically begin seeking efficiency by pruning less efficient and redundant flow paths, thereby reducing overhead. The effect is a streamlining of the system structure. Common system streamlining by organizations or programs during the development phase include reducing the number of suppliers to those offering the best value for money, bringing outsourced activities back in-house, reducing the number of products or services etc. Whether in social, political, or business environments the focus tends to be on repeating what has worked in the past. The process of replication implies standardization and a loss of flexibility. Progression from the growth phase through the development phase is typically slow and the transition from the growth phase to the development phase is usually more evident after the fact than during the transition. For systems displaying the familiar S-shape “growth” curves,

the transition from the growth phase to the development phase is usually prior to or in the vicinity of the inflection point in the growth curve. System dynamic models show that the inflection point for simple S-curve models having rapid feedback should be at the mid point between base level and the peak (or carrying capacity) of the S-curve. The slowing of growth is usually associated with either a peaking of resource availability and ability to repeat the successful patterns or to a saturation of the client base. Delays in the feedback process allow overshoot and either oscillation or collapse depending upon the nature of the critical resource.

The Maturation Phase

In the maturation phase the system growth slows further and often peaks as client saturation, resource limitations, competition, and other factors combine to restrict growth. Processes, procedures, and relationships will typically be highly standardized. Connections will have been pruned to the most efficient paths. Flow path alternatives of the system are streamlined and redundancy minimized. Overheads are minimal as efficiency is at its maximum. The major system elements and relationships become tightly linked along specific paths, leading to a fragility or brittleness of the system, making it more susceptible to disturbances and disruptions. Communication often becomes rigid as sensitivity to and awareness of potential sources of change is typically lost as environmental awareness wanes and past patterns are unquestionably taken as permanent.

Senescence

Events in senescence are highly dependent upon the nature of the system and its environment. In absence of a major disturbance or disruption, system throughput and activity may stabilize – ie the system will stabilize, cease to evolve, possibly continuing at the peak forming a plateau until something changes the environment. Eventually a disruption will usually stress a system beyond the level of possible accommodation by the highly efficient flow structure and the system will begin to decline. The speed of decline will vary with the nature of the disruption and the characteristics and fragility of the system. Frequently the system (family, nonprofit, corporation, etc) finds itself incapable of responding effectively to the change as flexibility of perception, thought, communication, and action have deteriorated in the rigidification associated with maturation. While generalizations about senescent systems are difficult, the behavior of a senescent system will range from stagnancy to death.

2. INSTABILITY

The normal progression of maturation described in the previous section is consistent with an environment that is reasonably stable and where major surprises are minimal. In the real world, the level of optimization (paring of inefficient connections) desirable from a long-term sustainability perspective is a function of the level of instability affecting the system. A mature system in an unstable world will need more connections and relationships (ie will be less

optimized) than a similar system in a more stable and predictable world. Less stable environments tend to keep the system “off balance” and thereby less efficient.

The key questions to consider initially are simply, “Has the environment been relatively stable or turbulent?” and “Does the level of turbulence seem to be increasing, decreasing, or remaining the same?” These answers become useful when the balance between maturity, connectivity, and instability are considered.

3. NETWORK CONNECTIVITY

The perspective of network connectivity is related to the overall pattern of connectivity of the system. Is it sparse, with only a few inputs per node or heavily cross-connected, with many inputs per node? Does one input dominate or are they relatively equal? Which input has the largest percentage of the total flow? Are the network of relationships and flows relatively linear? Or complex? Answering these questions in detail can require a great deal of time and effort, but one should begin by simply working with perceptions and feel. The following paragraphs provide background for framing the response to those questions.

Research into ecological systems provides insight into the balance between maturity, environmental instability, and the level of network complexity to be expected. In a stable environment (such as in a rainforest) specialization (sole sourcing) may be practical and highly linear systems may be viable (Ulanowicz 1997). However, as the level of instability increases, the ability to specialize decreases and more opportunistic strategies (with multiple sources of supply) would be expected to reduce the impacts of uncertainty. The ultimate insight from a wide range of network disciplines is that there should be between one and three effective sources for a network or system to be stable, ie the largest source will have a share between 33 and 100 per cent. Examples of stable and unstable network connectivity are illustrated in Figure 3 using the weight of the lines to represent the relative magnitude of the flows. At first glance this seems counter-intuitive.

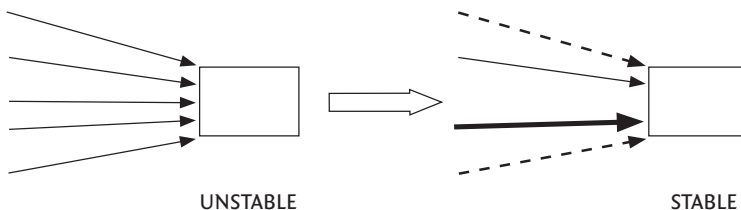


Figure 3: Examples of Unstable and Stable Network Configuration

The perspective of network connectivity also relates to the pattern of interconnectedness and interdependency of the system flows. During the growth phase systems frequently grow more complex as new relationships are explored. As maturity increases optimization tends to result in more linear, less interdependent flow structures as a result of paring of less efficient relationships. The ability to pare and streamline the system is limited by the instabilities. Mature systems will tend to maintain patterns of relationships over time. Mature systems are also likely to have feedback mechanisms, however, to help regulate flows in the system rather than the purely linear form shown below. Less mature systems will tend to display a more transient connection pattern as new relationships and combinations are tested and replace prior connections. Real systems should never look like the linear example in Figure 4. Expect some level of interdependence and feedback.

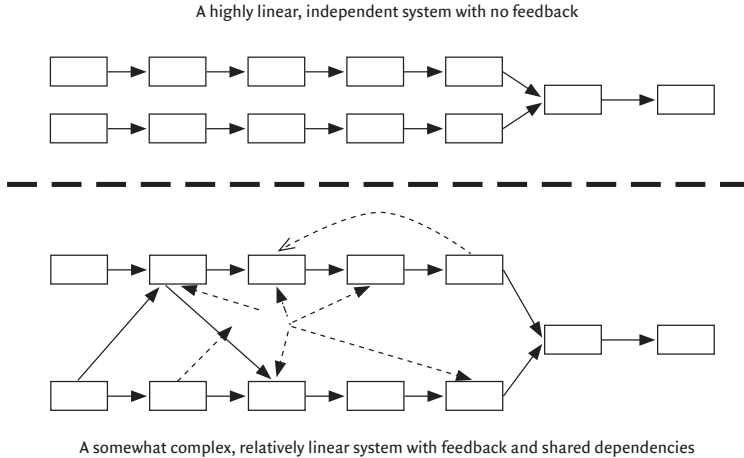


Figure 4: Linear and Complex System Configurations

It should also be noted that as systems grow more mature they tend to become more diverse – as a combination of optimization and specialization result in more nodes and interdependencies. However, the average number of inputs per node will tend to be less as specialization often implies that fewer inputs will be needed at each node.³

Given that background, one asks, does the system structure being evaluated seem consistent with the perceived maturity of the system and the perceived level of instability? If so, one moves on. If they do not seem consistent then it is

3 Think of a new organization where roles are still being defined and everybody is communicating with everyone else as compared to a mature organization where communication is strictly kept within the department except at the senior levels.

appropriate to dig a little deeper, discuss the perceptions some more and see if the discrepancies resolve. If the perceptions continue to be inconsistent, one can move on to other perspectives – but with questions overhanging the process.

Understanding Sources of Instability

The other five perspectives provide insight into potential sources of instability that could challenge the viability of the system being evaluated or stimulate changes in the system and its structure.

4. FITNESS COMPLEXITY

The perspective of fitness complexity arises from the study of mathematical biology and can be challenging. Fortunately one need not understand the underlying details to apply this perspective. The concept of fitness complexity is based upon the number of factors determining “fitness” in an evolutionary landscape. From a simpler perspective, this equates to how interconnected and interdependent the system is. Interpreting the implications can be summarized as two counterintuitive tendencies:⁴

Systems that are growing more interconnected and interdependent tend to:

1. Experience a “squeeze” to a plane – ie everyone tends to approach the average, as it grows increasingly difficult to maintain an advantage. Interdependencies limit the ability to separate one’s self from the masses. Those below average also tend to experience a “pull” toward the average as interdependencies aid them.
2. Become more turbulent and fragmented – with more specialties (special interest groups, specialists, etc)

While this perspective does not suggest details of “convergence” or how the environment will grow more turbulent, it does reinforce the concept that, for example, globalization will make it increasingly difficult for the United States or a leading manufacturer to maintain a unique advantage as globalization increases interdependence and provides a logic to support the rise of India and China as their fitness approaches the median. It further reinforces the logic that globalization will increase the number of “shocks” experienced as the impacts of formerly local events now ripple across the globe. At a smaller scale there is the tendency of community organizations to move towards “silos” of specialist agencies or develop patch wars between similar services.

From an evaluator’s viewpoint, the application of the fitness landscape perspective involves asking whether the system and its environment seems to be growing more, or less, interconnected and interdependent. If the environment is growing less interconnected then it is likely the environment will be growing more

4 Stuart Kauffman has written extensively on the topic of fitness landscapes and how they change with fitness complexity. (Kauffman 1993) Historically it has been assumed that globalization would lead to homogenization – and in some areas such as language it arguably has, but the clear, counterintuitive insight from this perspective is that increasing connectivity leads to fragmentation and specialization.

stable and predictable. The system may be able to optimize and increase fitness. However, if the environment is growing more interconnected, then it follows that the system should be expected to evolve in a manner that will be consistent with greater instability and that the participants in the system that hold advantaged positions will be under stress as their advantage declines.

The implications of this perspective span organizational, political and social realms. The Internet has enabled the creation of numerous listserves and Internet dialogues, representing a major increase in interconnectivity. That interaction enables the existence of special interest groups – ranging from sourdough bread to pedophilia. These interest groups occupy localized peaks in a fractured fitness landscape. While some of these groups are innocuous, others pose challenges to businesses, society, and governance.

This is, however, not a deterministic perspective. An evaluator facing growing complexity can recognize the implications and suggest programs reconfigure to better maintain advantages and meet future challenges.

5. SYSTEM BOUNDARIES

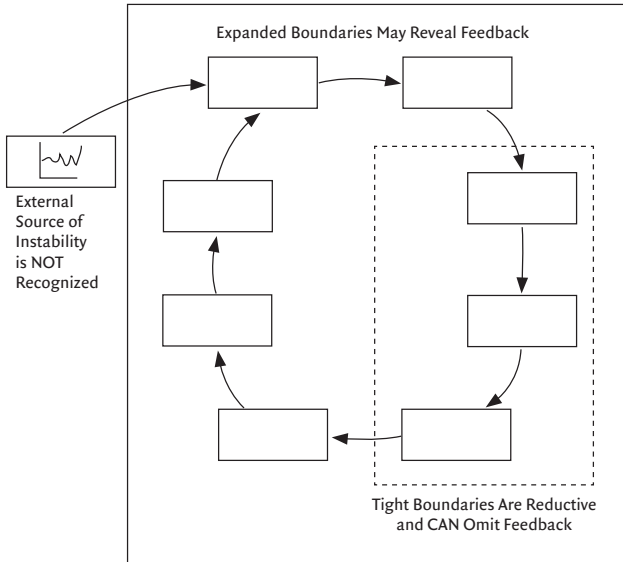
The setting of system boundaries has substantial influence on the perception of a system, and whether it is stable or turbulent. Suppose for example, you were standing in an ancient, mature forest where there is a constant profile of trees of a given age. The overall system is stable. However, at the local level, a tree dies and falls, creating a hole in the canopy. New trees and plants sprout and the local environment is immature and may take many years to blend into the surrounding forest. Figure 5 illustrates how boundaries may create a perception that a system is linear, or more complex, and may mask or reveal sources of instability. Examining an issue using tight boundaries such as those depicted by the dotted lines suggests linear causality and that no feedbacks will affect long term behavior. Expanding the boundaries to the solid lines introduces feedback with different behavioral implications as discussed in the next section. There will always be external, unrecognized sources of instability that lie outside the perceived boundaries of the system under study. Deciding what to include ultimately becomes a judgment call as one strives to include significant sources potential impact and exclude those that are trivial.

A useful approach involves building a causal loop or influence diagram describing the perceived system and scanning that diagram for factors that seem less stable or certain⁵. Exploring the boundaries of the model aids in recognizing potentially overlooked sources of instability so that one can proactively anticipate instability and potential solutions in advance. Trend extrapolation and cross impact analyses can also be useful in recognizing stocks that may serve as sources of instability.⁶

5 The process of building causal loop or influence diagrams is beyond the reach of this short article. See the chapters by Ken Meter, Richard Hummelbrunner and Dan Burke for additional information. The literature of system dynamics is also a good source of additional information.

6 Other chapters discuss the issue of boundaries, boundary setting and boundary critique. Martin Reynold's

Figure 5: System Boundaries And Perception



6. FEEDBACK STRUCTURES

Feedback structures are viewed as the ultimate long-term shapers of system behavior. Consequently recognizing feedback structures is critical in being able to recognize likely patterns of long-term system behavior. Feedback loops are routinely categorized as having positive feedback where an increase feeds back upon itself in a manner that stimulates further increases, and negative feedback where an increase stimulates reactions that tend to offset the increase and promote stability. It should be noted that positive feedback loops tend to encourage exponential growth and negative feedback loops tend to offset and limit growth from positive loops.

As feedback structures often shift in importance over time, it is important to consider not only historic and currently operating feedback structures, but also potential future structures. For example, successful programs and fads often begin with word-of-mouth structures driving the growth in numbers of clients and funders. Over time other factors can come into play, such as saturation of the market, loss of uniqueness, new competing products and services (or even service categories), etc. Exponential growth patterns resulting from positive feedback loops routinely transform into a “balanced” system as balancing (or

and Gerald Midgley’s chapters are particularly relevant.

negative feedback loops) come into play. System diagrams that show only positive feedback loops are invariably oversimplified. When building system diagrams it is appropriate to strive to recognize multiple balancing loops – loops that might be expected to counter any positive feedback loops. When exploring system boundaries it is always good to scan the system to see if inclusion of a new factor suggests possible feedback loops.⁷

As an aside which may make the concept of feedback loops more meaningful, when a client commissions an evaluation they are, in effect, initiating a new feedback loop that might be interpreted to be similar to that shown in Figure 6. In this diagram the issue being evaluated is shown in the box and is assumed to be an ongoing process or activity. The evaluation loop involves observation and assessment, communication to the client, and action by the client to modify the topic of study. While this simplified diagram glosses over complexities that often accompany studies, the intent of the loop is to improve the topic under study. It should also be noted that the loop may be interpreted as either a positive or negative feedback loop depending upon the impact of the changes. In either case, subsequent evaluation is appropriate, suggesting that evaluation should be a continual process.

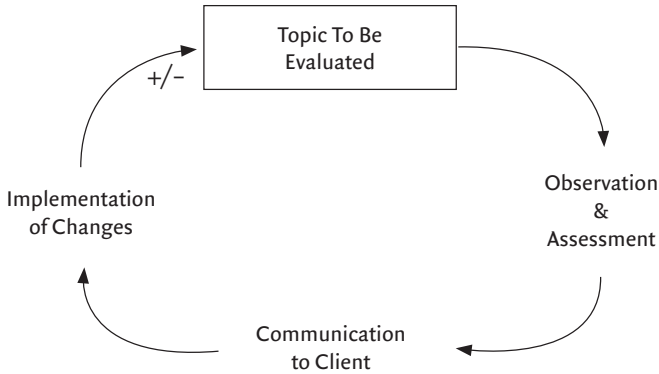


Figure 6: Evaluation as a Feedback Loop

7 A detailed treatise on feedback loops is far beyond the scope of this simple introduction. The concepts of feedback loops, and of positive and negative, or balancing, feedback loops are discussed extensively in the literature of system dynamics along with the implications of loops upon behavior. See also Ken Meter's, Richard Hummebrunner's, and Dan Burke's chapters.

7. STOCKS⁸

The concept of stocks and their dynamics in systems provide valuable insights to systems behavior. The concept that stocks drive systems is a fundamental tenet of system dynamics.

A brief review of the concept of stocks may be beneficial. Within the logic of system dynamics stocks are things that accumulate over time. Inherent to the concept of accumulation is that stocks are connected in a network of flows, both in and out. They may be tangible or intangible, real or virtual. Assets, skills, inventories, water (in a lake or bathtub), love, resources (such as political patronage), and clients are all examples of stocks. Stocks can be identified by the test of “freezing” time. If time stops, stocks continue to exist and have a level (ie amount of whatever it is the stock represents) while flows cease to exist. Newcomers to the concept of stocks will likely find that it takes some time to truly appreciate the power of the concept of stocks.

Stocks have important dynamic implications for a system. In most systems, stocks accumulate relatively slowly – though in a positive feedback loop they may seem to growing rapidly. By acting as accumulators stocks generally act in a manner that stabilizes the system to some degree (in absence of a catastrophe). The relative balance between a stock and its inflows and outflows can provide evaluators with insights into the ability of the stock to “dampen” disturbances, such as a large lake maintaining flow downstream during a drought, thereby dampening the impact of the drought. For instance, evaluating how well an agency drew on a stock of what initially appeared to be an unnecessary stock of unused but trained volunteers when funding for paid staff was withdrawn.

However, stocks can also reveal points of vulnerability of a system. Every stock is enabled by other stocks. For example, a lake can be thought of as being enabled by a watershed, an impermeable layer of soil or rock, and a dam. If anything happens to any of those stocks, the viability of the lake is placed in jeopardy. By inspecting the stocks (and implied stocks such as the watershed and dam for a lake) one can recognize potential sources of instability. Risk can be considered and the enabling stock can be included in the model if it is deemed to be a significant risk or an uncertainty deserving recognition. If the risk is adequate, appropriate actions could be taken to reduce the risk. Thus the concepts of stocks and flows provide evaluators with a base by which to judge a program’s readiness for shocks.

Stocks also offer an interesting approach to recognizing potential sources of future shifts in system structure. What is accumulating in the system under study? Does an industrial process have a substantial or difficult to dispose waste byproduct? Are an organization’s actions creating an accumulation of dissatisfaction with a program? Accumulations such as these often create a driver of change that ultimately leads to a shift in the system. Too often those impacted (the manufacturer or the program stakeholders) ignore the accumulations or fail

8 See Dan Burke’s chapter for an additional description of stocks and flows.

to recognize their potential until it is too late. Asking what is accumulating in the systems can be a productive approach to identifying potential sources of future system instability. It can help foresee “unanticipated” events – something of key interest to many evaluators and program providers.

8. STRUCTURAL DEPENDENCY

The final perspective of this methodology is structural dependency – ie the sharing of infrastructure by elements in the model. The importance of this perspective is particularly visible in supply chain issues but is valid for all systems. Multiple sources are independent only to the extent they do not share infrastructure.

Having two separate projects provide sex education programs to schools may help a Principal negotiate the tricky politics of that topic, but not if both projects are funded by the same Foundation that decides, like many do, to change its focus every few years. For maximum robustness shared structural dependencies should be minimized to the extent practical. Shared dependencies identify points of increased risk to the system.

Summary

Experience with this methodology suggests the logic is particularly useful for gaining insight into a system prior to using other, more detailed and reductive methodologies. The methodology is also useful when systems are highly complex and causal mechanisms are controversial. It also accommodates situations in which the structure is known to be in flux.

The process is flexible and offers opportunity for blending with other approaches as is appropriate to the study at hand. Successful applications of this logic have ranged from quick exploratory system reviews to cyclical, formal, detailed assessments. In any event, recognition of the systems insights from this methodology seems to encourage better appreciation of system characteristics, tendencies, and potential behavior.

Developing fluency in this logic requires time and experience. Additional readings are available at the author’s website: <http://www.jayforrest.com>.

Bibliography

- Forrest, C.J. 2004. *Evolution and Behavior of System Structure: Eight Perspectives for Examining Complex Issues*. Paper presented at the 22nd International Conference of the System Dynamics Society, Oxford, England.
- Kauffman, S. 1993. *The Origins of Order*. New York: Oxford University Press.
- Pimm, S L. 1982. *Food Webs*. London: Chapman and Hall.
- Ulanowicz, R.E. 1997. *Ecology, The Ascendent Perspective*. New York: Columbia University Press.

CONCLUDING COMMENTS

Iraj Imam, Amy LaGoy, Bob Williams

Now what? How to promote systems concepts in evaluation.

Between the new language and new concepts in this volume, evaluators will have found much that is familiar. There are many similarities between evaluation and systems approaches. Where differences provide opportunities for learning, similarities provide the solid ground on which systems and evaluation practitioners can build bridges, share, learn from each other, and develop. We expect that evaluators may already share with systems thinkers certain methodological attitudes and approaches, particular beliefs, and orientations to data, to people, to problems, and situations. We describe some of these below.

Certain established evaluation methodologies fit well inside systems concepts – the various forms of “Democratic Evaluation” are perhaps the most obvious. Other approaches such as Program Theory, Empowerment Evaluation, Utilization-focused Evaluation, and Realistic Evaluation share a portion of common ground. Both fields have also been influenced by action research concepts and methods.

Many evaluators understand programs as arrangements of interacting policies, practices, people, and places. Evaluators know that a change in one of these elements is likely to lead to change in another; that some changes will have marked effects, while others will not.

Generally systems people and evaluators understand that programs help shape and are shaped by the worlds in which they operate and do so in ways that question our assumptions about linear causal relationships (eg the giraffe and tree example in Gerald Midgley’s chapter). They both also know that situating a program in its socio-cultural, historical, economic, political, and organizational context is important to understanding how and why the program works as it does.

Members of both fields explore a program’s intended and unintended consequences. Yet, we all realize that we simply cannot pursue all lines of inquiry and bound the scope of our study. Evaluators do this by determining—often with some knowledge of a program’s theoretical underpinnings and with the help of program stakeholders— which elements of the program and its context are directly relevant to their inquiry and which fall outside its parameters. Systems people do much the same, although they have special and arguably more rigorous ways of deciding where and how to set boundaries.

Though evaluators often find themselves monitoring the extent to which programs meet their stated goals, many share the systems field’s interests in improving the state of things. Both are concerned with what can be, and how it might become that. Evaluators often work with clients and other program stakeholders to ensure that among the multiple perspectives considered in the program evaluation are those of the less or least powerful, the marginalized.

Evaluators also take care to be alert to the ways that our own attitudes, biases, proclivities, and place in the world shape how we read and interact with a program.

Finally, as several authors mentioned, evaluation can be seen both as a system itself, and as a sub-system that provides feedback to a broader system.

With all these similarities and connections it would be nice at this point to provide a general taxonomy of systems and evaluation methods – in evaluation situation “X” use systems method “2”. During the production of this volume the authors and editors made several attempts to construct such a taxonomy. We know others are also working on this; most notably Derek Cabrera & Bill Trochim (2006) as well as authors Boon Hou Tay and Martin Reynolds. So why no neat table ?”

Firstly, neither the systems nor the evaluation fields have generally accepted taxonomies of their own. Attempts within the systems field (eg Jackson and Keys 1984) have been criticized for being oversimplistic, and Dan Stufflebeam’s attempt within the evaluation field – his evaluation models – has suffered the same fate (Stufflebeam 2001). Without accepted taxonomies within each field, a framework that joins the two fields is a bit like pinning a tail to a tail – there would be no substantial body to support it from either field.

Secondly, it may be too early. Taxonomies have a habit of creating the impression of certainty where there is only probability. This is very early in the relationship between systems and evaluation – let’s see how the relationship develops in practice for a couple of years before beginning to codify it.

That period of uncertainty may disturb some evaluators and evaluation clients, but it provides a creative opportunity for those willing to experiment and try things out. Over time a record of success and failure will provide a body of knowledge that will help braid the two fields with increasing certainty This volume is a step in that direction – and there will be other increasingly sure steps as we adapt, adopt, and learn.

But we can stick our neck out a little. Based on current thinking (eg Cabrera & Trochim 2006), discussion between this publication’s authors, and gazing into our crystal ball we can begin to predict what might be the dimensions that will form the basis of future taxonomies linking systems concepts with particular evaluation situations. These include :

- The nature of the evaluation situation (eg disputed, multi-purposed)
- The state of the situation (eg stable, volatile, mature, life cycle stage)
- The scale of the situation (eg big, small, both)
- The structure of the situation (eg simple, complicated)
- The perspectives on the situation (eg multiple, single)
- The nature of the evaluation questions (eg open, closed, deterministic, exploratory)

We also predict that evaluators will not necessarily use entire systems methodologies (eg cybernetics, soft systems) but take specific methods from systems methodologies (eg viable systems model, CDE, CATWOE, critical system

heuristic, CLDs). There are of course risks with that, but with care they are not insurmountable – as many of the chapters in this volume illustrate. Indeed this may be a good place for you to pause and consider what each chapter has contributed to this emerging taxonomy.

In conclusion

We believe that the use of systems concepts and approaches can significantly improve the relevance and utility of evaluation. It can do this by helping stakeholders clarify their respective interests and power and the worldviews implicit in their program. It can help clarify the goals, roles, responsibilities, and knowledge requirements of an evaluation. It can be used at the design, implementation, analysis, and reporting stages of an evaluation. Systems concepts and approaches can be mixed and matched according to the circumstances.

Systems approaches can be powerful means for evaluation participants to see the world in many different ways (stepping into each others' shoes, so to speak) increasing their insights and potentially, their responsiveness to ideas and attitudes different from their own. Evaluators can tap this knowledge to design a more relevant evaluation, increase participation in the evaluation process, and enhance the usefulness of evaluation findings. Most systems approaches do not prescribe what to do with the increased insight and creativity. So evaluators and other stakeholders will have to decide how to use this deepened understanding in the evaluation process and beyond.¹

The twelve chapters in this volume show how all this is possible and practical. We carefully selected contributors who are systems practitioners working in the evaluation field. Many more examples emerged as we worked on this volume, so we know that systems approaches can be used in evaluation work. We know that systems concepts and methods offer evaluators additional, valuable, ways of framing and exploring the complex tasks they face. The difference between the two fields is important – indeed essential, but the common ground between the two fields is broad enough to build on.

We hope you will help with the construction. Gerald Midgley encourages us to start where we currently are and build on that. In the wide array of possibilities the chapters have laid before you there will be an idea, a concept a method or a methodology that provides the tools for you to begin the construction. Good luck, have fun, and as on all building sites keep an eye out for falling debris !

1 Our thanks to Dan Burke for these insights.

References

- Cabrera, D and Trochim, W. 2006. A Protocol of Systems Evaluation. In, D Cabrera (ed), *Systems Evaluation and Evaluation Systems Whitepaper Series*. Ithaca NY: Cornell University National Science Foundation Systems Evaluation Grant No. EREC-0535492.
- Jackson, M C and Keys, P. 1984. Towards a system of systems methodologies. *Journal of the Operational Research Society*, 35: 473-486.
- Stufflebeam, D L. 2001. Evaluation Models. *New Directions for Evaluation*, 89: Spring 2001. San Francisco CA: Jossey-Bass.

BIOGRAPHIES

KATE ATTENBOROUGH

Yemeni Economic & Training Centre, Sheffield, South Yorkshire, England
kattenborough@yetc.co.uk
+44 0114 261 8620

Kate is a practitioner, using soft systems methodology in the public sector for evaluating and designing training interventions, organizational processes and programmes, and in assessing diverse and sometimes contradictory findings of labour market research. She has since use the approach extensively in the voluntary and community sector, as both manager and freelance consultant, finding the methodology invaluable in assessing and dealing with complex situations, projects and community politics. Currently a volunteer management teacher with the Yemeni Centre, her most recent use of the methodology was the catalyst that created a partnership with Sheffield Hallam University to uniquely deliver an initial teacher training course in the community, bringing in bilingual teachers from differing cultures.

She is at present exploring the use of soft systems in adult education, in the areas of curriculum design and development, research and evaluation.

RICHARD BAWDEN

Visiting Distinguished University Professor, Michigan State University,
East Lansing, Michigan

Richard Bawden has been swimming in the system's stream for more than four decades in his struggles with the evolution of processes of, and strategies for, responsible rural development that are inclusive of people and nature alike. For a long while (1978–1993) he was Dean of the Agriculture and Rural Development team at Hawkesbury Agricultural College University of Western Sydney. During this time, his essential efforts, in close collaboration with his academic colleagues, students, and a wide variety of stakeholders, were focused on providing formal experiential education opportunities for the development of “systemic competencies” by all involved in the search for “rural betterment”. For the past ten years or so – the majority of which he has spent as a Visiting Distinguished University Professor at Michigan State University – this focus has been extended to embrace the broader question of the manner by which universities as institutions can ‘engage’ with the citizenry and with other relevant social institutions, in the quest for, and use of, systemic approaches appropriate to the development process within a wide spectrum of domains and across a diverse range of national cultures. Evaluation has been a formal and importantly explicit dimension throughout all of this work.

DANIEL D BURKE, Ph.D.

former Deputy Director for Education
The CNA Corporation
Alexandria, VA

Dr Burke has studied the implementation of systemic reform in K-12 education while working with the 25 largest school districts in the United States. He has done research to identify and understand the critical factors in recruiting and retaining high-quality m/s teachers. His work in K-12 education began with the design of professional development programs in general science for K-8 teachers and in molecular biology for high school teachers. His knowledge of the organization and function of K-12 education systems was gained through his service as Senior Staff Associate for Systemic Reform, Directorate for Education and Human Resources, NSF.

Presently, Dr Burke studies the use of system dynamics approaches and computer modeling to evaluate the effectiveness of educational systems. He has developed computer simulations to examine critical factors in professional development, in systemic reform of K-12 education, and a model tracking the flow of minority students from entering college to becoming tenured faculty and in the recruiting and retention of science and mathematics teachers.

DANNY BURNS

Professor of Social & Organisational Learning
SOLAR, University of the West of England
danny.burns@uwe.ac.uk
www.uwe.ac.uk/solar
+44 117 328 1113

Prior to current position Danny worked at the School for Policy Studies, University of Bristol, where he directed the M.Sc. in Management Development and Social Responsibility. In the early 1990's Danny was Director of the Tenant Participation Advisory Service for Scotland and prior to that Director of the Decentralisation Research and Information Centre. Danny's work focuses on community development and participation, local governance, and participatory learning. His most recent book is *Community Self Help* (Palgrave) and he is currently completing a book on Large System Action Research that will be published by Policy Press.

GLEND A EOYANG

Executive Director
Human Systems Dynamics Institute
geoyang@hsdinstitute.org
www.hsdinstitute.org

Glenda Eoyang is founder of the Human Systems Dynamics Institute, a research and consulting group developing theory and practice in the emerging field at the intersection of complexity and social sciences. She began her work with complex systems in 1989 and received the first doctorate in Human Systems Dynamics from Union Institute and University in 2002.

Eoyang's theoretical work covers a range of models and approaches. She used nonlinear time series modeling to investigate marketing strategies, created a computer simulation model that functions as an executive decision support tool, designed a health care simulation game, and discovered a fundamental model for the conditions for self-organizing in human systems (CDE Model). In addition, Eoyang teaches, speaks publicly, and practices the principles of complexity in business, industry, nonprofit, and government settings. She has written numerous articles for academic and business publications on topics ranging from fractals for business administration to human computer interface design, youth gangs, productivity, large group events, team building, sustainability of organizational change, and program evaluation. Her first book, *Coping with Chaos: Seven Simple Tools* (Lagumo, 1998), is an accessible and useful treatment of human systems dynamics for first-line managers and supervisors.

DALE FITCH

Assistant Professor
School of Social Work
University of Michigan 1080 S. University, Ann Arbor, MI 48169.
dale@umich.edu

Dale Fitch received a Master of Social Work (1984) and a Ph.D. in Social Work (2001) from the University of Texas at Arlington. He is currently an Assistant Professor in the School of Social Work at the University of Michigan. Dale has a broad range of social work experience including private practice, institutional mental health, and hospital social work. He uses both quantitative and qualitative research methodologies addressing topics that include management of information systems in human service organizations, systems theory, and decision-making. Dale has practice interests in human services administration, community practice, and policy. Recent projects involve the design and development of web-based management information system that facilitates inter-organizational case management.

JAY FORREST

San Antonio, Texas

systems@jayforrest.com

<http://www.jayforrest.com>

Jay is currently an aging doctoral student in Foresight and Futures Studies at Leeds Metropolitan University in the UK. His research strives to apply cutting edge systems theory to futures studies to build a better understanding of the uncertainties that shape alternative futures. At the same time, Jay is striving to build a systems logic that facilitates recognition of uncertainties, providing a stronger basis for bounding models, problems, and issues with the ultimate objective of better strategies for proactive action.

Jay became interested in system dynamics and systems thinking while working with clients on business problems in the 1980s. Over time he became intrigued by “why models fail” and began seeking ways to build better models. While obtaining a masters degree in Futures Studies from the University of Houston at Clear Lake, he learned to embrace uncertainty and began exploring methods of blending systems thinking and futures studies that would strengthen both fields. Jay is a founding member of the Association of Professional Futurists, and is a member of the System Dynamics Society, the International Society for the System Sciences, and Institute of Electrical and Electronics Engineers. He is also active in the Communities of the Future movement. Additional papers and information are available at his web site.

RICHARD HUMMELBRUNNER

Senior Associate

ÖAR Regionalberatung

Alberstrasse 10, 8010 Graz, Austria

hummelbrunner@oear.at

<http://www.oear.at>

Richard Hummelbrunner originally trained as an economist and spatial planner and has acquired more than 30 years of professional experience as consultant in the field of local and regional development. Faced with the challenges of increasing complexity in his work, he became interested in the systems field in the early 1990s. He then trained extensively in systemic intervention and advisory techniques and began exploring the use of systems thinking in regional development.

Nowadays he predominantly works as evaluator, advocating and developing evaluation styles which incorporate concepts and techniques derived from systems thinking. He is or has been team leader of several major evaluation assignments (e.g. various Structural Fund Programmes in Austria, INTERREG and LEADER Programmes) and has also taken part in evaluation assignments at transnational and European level.

IRAJ IMAM

Center for Applied Local Research (CAL Research)

5200 Huntington Ave. Suite 200

Richmond CA 94804

United States

iimam@cal-research.org

<http://www.cal-research.org>

Iraj Imam studied architecture in National University in Tehran, where he made documentary movies exploring the Persian concept of architectural, as well as social, space. In 1974, he came to the US from Iran to study sociology of development. At UCLA, he completed his Ph.D. on geographical uneven development in petroleum-based economies (1985). He conducted research at both UCLA and UC Berkeley on paradoxes of social and spatial uneven development in oil-based countries. He joined The Center for Applied Local Research (CAL Research; a non-profit consulting organization) in 1993 and has managed local evaluation projects in substance abuse and mental health prevention and treatment, youth development, and criminal justice areas. As a Senior Research Manager at CAL Research, he is currently working with service providers to developing and applying spatial theory concepts and systems approaches to evaluation. Given the current re-emergence of reductionist worldviews and the new aggressive governments, he is interested in developing critical and transformative ideas and practices that address social and power inequalities in evaluation and in human services.

AMY LA GOY

Amy LaGoy, a consultant in Berkeley, CA, has been evaluating education and public health programs since the late 1980s. Although her evaluation work typically examines the alignment of a program's outcomes with its stated goals and objectives, Amy also works with clients to identify unanticipated program impacts and to explore why a program works the way it does. These latter two strands of evaluation inquiry – and the challenges embedded in them – primed Amy to learn about systems thinking and to consider its relevance for evaluation work.

In the early 2000s, Amy oversaw the evaluation division of a Berkeley consulting firm. She has taught graduate seminars on program evaluation, research methodology, and qualitative data reduction and analysis at U.C. Berkeley and San Jose State University. She earned her doctorate in Education at the University of California, Berkeley. Her dissertation, an ethnographic study of the academic and social worlds of 71 high-school students, explored the interaction of gender, students' social status, and their academic experiences. She has an M.A. in Educational Administration from U.C. Berkeley and a B.A. in English from Mount Holyoke College in Massachusetts. After college, Amy went to Yemen to teach English with the American Peace Corps.

BOBBY, KEE PONG, LIM

Manager, Rail Operations Training
SBS Transit
Singapore

Bobby Lim has been working in the area of rail operations and trainings for more than 20 years. In his present job as the Manager for Rail Operations Training in SBS Transit, he is responsible for leading a team of training specialists and operations trainers in the design, development and implementation of a full spectrum of operations training courses needed by SBS Transit in order to meet the existing and emerging training needs of traffic controllers, station managers, assistant station managers and customer service officers.

Prior to joining SBS Transit, Bobby Lim has 14 years of working experiences with the operations department of the Singapore Mass Rapid Transport (SMRT). SMRT is another main railway service provider in Singapore that has empowered Bobby Lim with the in-depth knowledge of railway services. Bobby Lim holds a graduate diploma in Personnel Management with Distinction, diploma in Human Resource Development and a diploma in Civil Engineering.

KEN METER

President, Crossroads Resource Center
7415 Humboldt Ave. S.
Minneapolis, Minnesota 55423
+1 (612) 869-8664
kmeter@crcworks.org
www.crcworks.org

Ken has worked alongside rural leaders in local economic analysis, worked alongside inner-city residents as an evaluator for two long-term poverty-reduction efforts in California and Minnesota, and directed community process and indicator selection for the city of Minneapolis Sustainability Initiative. He has taught economics in the Department of Applied Economics at the University of Minnesota, and at the Kennedy School of Government at Harvard University, where he studied as a mid-career Public Service Fellow.

GERALD MIDGEY

Senior Science Leader at the Institute of Environmental Science and Research (ESR), Institute of Environmental Science and Research (ESR), 27 Creyke Road, PO Box 29-181, Christchurch, New Zealand.

Adjunct Professor in the School of Management at Victoria University of Wellington, New Zealand.

Centre for Systems Studies, Business School, University of Hull, Hull, HU6 7RX, UK

Gerald has undertaken a wide range of systemic evaluations and interventions commissioned by public and voluntary sector organisations, and his engagement in practice has directly informed his theoretical and methodological writings. He has also had over 200 papers published in international journals, edited books and practitioner magazines, and has written or edited nine books. These include: *Systemic Intervention: Philosophy, Methodology, and Practice* (Kluwer, 2000); *Systems Thinking, Volumes I–IV* (Sage, 2003) and *Community Operational Research: OR and Systems Thinking for Community Development* (Kluwer, 2004).

MARTIN REYNOLDS

Lecturer

Systems Department, The Open University, Milton Keynes, MK7 6AA, UK.

m.d.reynolds@open.ac.uk

<http://systems.open.ac.uk/page.cfm?pageid=MartinRhome>

Martin is a lecturer and researcher in the Systems Department at The Open University. His interest in systems thinking began whilst evaluating participatory natural resource-use appraisal as part of doctorate studies at the Institute for Development Policy and Management in Manchester, UK. The failure of participatory development in less-developed countries like Botswana, despite considerable financial and expert support, triggered an interest in the potential synergies between development practice and critical systems thinking. Martin's work drew on prior practical experience of working in the village of Mochudi in Botswana (original home of Precious Ramotswa, from the *No 1 Ladies' Detective Agency!*) between 1983 to 1992.

Before joining The Open University in 2000, Martin worked with Gerald Midgley on an action research programme involving environmental planners and systems/ operational researchers in the UK. At the time of writing, Martin is working with Professor Werner Ulrich from Switzerland in piloting formative natural resource-use evaluation with stakeholders in Guyana, including representatives of the Amerindian tribe.

BOON HOU, TAY, PhD

Director and Project Manager of IN Technology Pte Ltd.
Singapore

Boon Hou pioneers three strategic research activities in the company, namely, Artificial Intelligence, Systems engineering and Datalink. The first involves design and development of an intelligent diagnostic expert system shell (DES) for implementing expert systems and embedded solutions in automotive industries. The second involves the use of Systems Engineering Concept, a logical sequence of activities and decisions that transforms an operational need into a description of system performance parameters and a preferred system configuration. The third focuses on wireless communications.

Dr Tay received his Bachelor in Electrical and Electronic Engineering (Hons) in Telecommunications from Queen Mary, University of London in United Kingdom. He also won the Drapers Scholarship from Queen Mary in 1986. He received his Doctor of Philosophy from Graduate College of Management, Southern Cross University in Lismore, Australia. He is the Singapore Industrial Representative for the Technical Publications Specification Maintenance Group (TPSMG) that is responsible for maintaining the S1000D specification for the European Association of Aerospace Industries (AECMA). He is a Member of the Systems Engineering Society of Australia.

BOB WILLIAMS

22 Rakeiora Grove, Korokoro, Petone, Wellington, New Zealand
<http://users.actrix.co.nz/bobwill>
bobwill@actrix.co.nz
+64 4 586 2790

Bob has been at the forefront of promoting systems based approaches in evaluation. He has conducted workshops for evaluators on this topic in New Zealand, the UK, the USA, Canada, and Australia. He has considerable experience writing and talking on systems, action research and evaluation topics in journals, books, Internet discussion groups and conferences.

Originally trained as an ecologist, Bob has worked as an environmental and social researcher, community worker, policy analyst, boat cleaner, bus conductor, and electricians mate. He left Britain the late 80s, sick to death of what Margaret Thatcher had done to the place, and now lives in Wellington, New Zealand. He is a qualified snowboard instructor – and spends the winter months mostly doing that.