

Evaluating Complexity: Theory, Evidence and Practical Guidance

Professor Ian Sanderson

Policy Research Institute, Leeds Metropolitan University

**Paper prepared for Symposium on ‘Theory-Based Evaluation’,
European Evaluation Society: Fifth Conference, *Three Movements in Contemporary
Evaluation: Learning, Theory and Evidence*,
Seville, Spain, October 10-12th, 2002**

This is a draft in need of further work. Please do not quote without permission. Comments welcomed.

**Professor Ian Sanderson
Policy Research Institute
Leeds Metropolitan University
Bronte Hall
Beckett Park Campus
LEEDS
LS6 3QS**

**Tel: 0113 283 1747
Fax: 0113 283 1748
Email: I.Sanderson@lmu.ac.uk**

Evaluating Complexity: Theory, Evidence and Practical Guidance

**Professor Ian Sanderson,
Policy Research Institute, Leeds Metropolitan University, UK**

Abstract

Recent work on complexity theory has significant implications for our conceptions of the nature of social systems and for ideas about how we can gain an understanding and achieve beneficial change of such systems. This paper seeks to identify some of the key implications for our thinking about the role, purpose and practice of evaluation of policy initiatives in complex, dynamic social systems, with a particular focus on theory-based evaluation. It covers issues relating to how we can gain an understanding of complex systems, how we can change social systems through policy action, and how we can evaluate such policy action to achieve understanding, learning and practical guidance.

Introduction

The contemporary pace and character of technological and social change has generated a climate of uncertainty in which traditional ways of thinking about social systems and processes of social change have been fundamentally challenged. For postmodernists the ‘messiness of the human predicament’ implies an end to rationalist attempts to ground policy action on evidence-based learning founded upon theoretical knowledge. An important aspect of the rationalist response is the search for a middle ground between ‘foundationalism’ and ‘anarchic relativism’ that provides an adequate evidential basis for action to address social problems. However, there is intense debate in the philosophy of social science about what grounds can provide a legitimate warrant for evidence.

The resurgence of interest on theory-based evaluation reflects this rationalist response but also embodies epistemological dispute. At the base of the dispute are differing assumptions about the nature of the social world and how we can gain an understanding of it. These assumptions in turn underpin conceptions of how social systems work and how we can change them to produce more ‘socially desired’ outcomes. Such conceptions then influence our view of the rationale and purpose of evaluation, and of how evaluation can derive evidence to improve our policy efforts.

The purpose of this paper is to consider the implications of recent thinking about complexity for these assumptions and conceptions about social systems and the role of evaluation and, specifically, the implications in relation to theory-based evaluation. Work on non-linear, open, dynamic, non-equilibrium systems has undermined many of the key assumptions that have underpinned traditional social and economic theories and research methodologies. From the perspective of complexity theory, can we retain an optimism about our capacity to understand and change social systems? What are the implications for our notions of ‘theory’ and ‘evidence’ and approaches to evaluation? And, more broadly, do we need to change our conception of what constitutes ‘rationality’ in addressing the ‘messy human predicament’?

The paper is structured as follows. The next section provides a very brief overview of the key notions of complexity theory and their roots in work on chaos. This is followed by a discussion of the implications of these ideas for traditional notions of rationality underpinning social science. The next section considers the implications for our thinking about our capacity to plan and change social systems through governance processes and this is followed by a critical look at our ideas about the role of evaluation in evidence-based approaches to policy making, particularly, the role of theory-based evaluation. This section ends by revisiting prevailing notions of rationality underpinning our thinking about policy making and evaluation. The final section provides a very brief consideration of some aspects of evaluation practice and the implications that derive from work on complexity.

Chaos and complexity

Over recent years, scientists have increasingly acknowledged that the world “...is no longer symbolized by the stable and periodic planetary motions that are at the heart of classical mechanics. It is a world of instability and fluctuations, which are ultimately responsible for the amazing variety and richness of the forms and structures we see in nature around us” (Nicolis and Prigogine, 1989, p.ix). It is recognised that new concepts and methodological tools are necessary to analyse complexity “...in which evolution and pluralism become the key words” (ibid.).

Nicolis and Prigogine (ibid.) stress the key role of two disciplines. Nonequilibrium physics deals with the properties of matter in far-from-equilibrium conditions while the modern theory of dynamical systems emphasises the prevalence of instability whereby small changes in initial conditions may lead to large amplifications of the effects of the changes. This phenomenon, known as ‘sensitive dependence on initial conditions’, produces the key characteristic of dynamical systems – unstable, aperiodic behaviour that is intrinsically unpredictable (Kellert, 1993, p.x).

Notions of complexity have their roots in chaos theory:

Chaos theory is the result of natural scientists’ discoveries in the field of nonlinear dynamics. Nonlinear dynamics is the study of the temporal evolution of nonlinear systems. Nonlinear systems reveal dynamical behavior such that the relationships between variables are unstable. Furthermore, changes in these relationships are subject to positive feedback in which changes are amplified, breaking up existing structures and behavior and creating unexpected outcomes in the generation of new structure and behavior. These changes may result in new forms of equilibrium; novel forms of increasing complexity; or even temporal behavior that appears random and devoid of order, the state of ‘chaos’ in which uncertainty dominates and predictability breaks down (Elliot and Kiel, op. cit.: 1).

Kellert (op. cit., p.2) defines chaos theory as “...the qualitative study of unstable aperiodic behaviour in deterministic dynamical systems.” Key characteristics of dynamical systems, then, are instability and aperiodic behaviour. Unstable behaviour means that the system never settles into a form of behaviour that resists small disturbances while aperiodic behaviour occurs when no variable describing the state of the system undergoes a regular repetition of values. Therefore the behaviour of such systems is highly complex and exact predictions are impossible (ibid., p.4).

Marion (op. cit.: 6-7) argues that while chaos theory is more suited to natural systems, notions of complexity can take account of “...such issues as adaptation, deliberative behavior, intelligent behavior, reproduction and evolution...” in social systems and allow some degree of stability and inertia. Reed and Harvey (op. cit.: 359) argue that complex systems “...defy standard positivist canons of description, prediction and explanation...”; their behaviour results from dynamic interaction between their component parts over time and emerges as the ‘holistic sum’ of these interactions. Functionalist notions of order-seeking, near-to-equilibrium social arrangements are inappropriate; rather, it is argued, the principles of ‘dissipative systems’ are applicable to the analysis of social life.

Dissipative systems are the subject of study in non-linear thermodynamics (Prigogine, 1996). Because such systems are thermodynamically open, they are capable of assimilating large amounts of energy from their environments and converting this into increasing structural complexity. This process is irreversible so dissipative systems possess an evolutionary capacity (based on so-called 'negentropic' processes) which enables them to avoid reversion to a state of thermodynamic equilibrium. Dissipative systems which can sustain these processes supporting the evolution of increasing structural complexity can, in certain circumstances, undergo sudden, rapid, transformational change (a so-called 'bifurcation') and evolve to a new state. The achievement of such transformational change is very sensitive to the appropriate conjunction of 'initial conditions' and is therefore non-replicable over time; knowledge of a system's behaviour in the past will provide little guide to likely future behaviour (Harvey and Reed, op. cit.: 320-5; Reed and Harvey, op. cit.: 361-5; Elliot and Kiel, op. cit.: 5-6).

The problems with prediction, however, do not imply that we cannot explain the behaviour of complex dynamical systems; according to Reed and Harvey (ibid.: 364) "...we can at least reconstruct the particular constellation of structured choice and accident that led to the present reality." But approaches founded upon the assumptions of stability and equilibrium, of linearity in the relationship between variables, and of proportionality of change in response to causal influences are not appropriate in seeking to understand social systems that exhibit complexity.

Complexity and the 'rationality project'

The implications of complexity are far reaching for traditional approaches to social science. The implications at the philosophical level are considered by Rescher (1998) who argues that "...the idea of complexity is effectively absent from most metaphysical systems" (ibid., p.7):

Ever since the origination of the classical idea of a 'science' in Greek antiquity, most philosophers have geared their aspirations for the discipline to the idea of a science modelled – as was the science of the Greeks – upon *mathematics*. Their goal has been to answer the questions of their field with the precision and universality characteristic of the mathematical sciences. The theories of such a 'science', as classically conceived, seek to accomplish their explanatory business in terms of an unqualified universality based on what happens always and everywhere and in all circumstances. The quest for a cognitive grasp of the lawful structure of the world in terms of a deep irrefrangible necessity underlying the world's seemingly fortuitous phenomenal surface has ever been the aspiration of science (ibid., p202-3).

For Toulmin (2001), the problem lies in the seventeenth century 'quest for certainty' under the influence of Descartes' work on "...a universal system of physics expressed in mathematical form...", which culminated in Newtonian dynamics being taken as the exemplar of rationality (ibid., p.31-48). Thus, from this time, "...the ideals of rational intelligibility and intellectual order current in Europe emphasised regularity, uniformity, and above all stability" (ibid., p.48).

However, as Dupre (2001, p.165) argues, "...the ability to apply Newtonian laws to more complex systems has proved severely limited..." and "...far from knowing that these laws are universally true, we know that they are generally false" (ibid., p.166). Moreover, Toulmin

(op. cit., p.47) suggests that the emulation of Newtonian dynamics by the social sciences “...as the type example of a Serious Science...” in the quest for prediction and control was misguided since these goals had never been achieved even in planetary astronomy (ibid., p.48-54). The social sciences thus pursued a vision of a ‘Physics that Never Was’.

Nevertheless, the dominance of this paradigm of stability, linearity and regularity – supported, Kellert (op. cit., p.154-5) argues, by a cultural and economic interest in prediction and control – effectively suppressed the study of non-linear dynamical systems and complexity. The recognition that “...instability and disorder are not only widespread in nature, but essential to the evolution of complexity in the universe...” (Elliot and Kiel, 1996, p.2) has profound implications. Rescher (op. cit., p. 203-4) argues that it calls for a more modest, empirically-oriented programme of philosophical inquiry, which forsakes the search for universal laws for a less ambitious concern with ‘limited and contingent generalities’, “...guided by the teachings of our experience with the world’s course of affairs.” He summarises the implications of such a ‘phenomenological turn’ in philosophy as follows:

(P)hilosophical standardism opts for practicable modesty in a complex world. It calls for abandoning pretensions to universality and inevitability, instead shifting our attention to normal circumstances and ordinary cases. Gearing its philosophizing not to the necessities of abstract general principles, but to the actualities as our cognitive commerce with the world reveals them to us, philosophical standardism abandons the traditional a priori stance of mainstream Western philosophy in favor of one that is a posteriori and experientially oriented. It does not try to resolve issues in terms of how things must necessarily go as a matter of theoretical general principle, but rather proceeds in terms of how things do normally go as best we can determine the matter. Rather than looking to general principles of abstract necessity, it proceeds in an empirical spirit, seeking guidance in *experience* – with all its characteristic contingencies and potential incompleteness (ibid., p.204).

This position has echoes in Dupre’s (op. cit.) call for pluralism in scientific reasoning that recognises the need for “...a more synoptic and integrative vision than the analytical methods of science allow” (ibid., p.185), a vision that accommodates ‘practical, experiential wisdom’. In his call for a restoration of the ‘Balance of Reason’, Toulmin (op. cit.) emphasises the importance of practical wisdom and tacit knowledge based upon experience in making ‘reasonable judgements’. Again, this requires a pragmatic modesty and pluralism:

If a representation pulls the phenomena together in a way that makes intelligible sense, giving a systematic grasp of the field concerned, that is enough ... no kind of explanatory representation can suit all kinds of phenomena, but one or another model has practical results that meet the needs of some particular natural or human science, and the empirical limits of the model are thus explicitly recognised. This accomplishes, all kinds of thought and action that engage our Rationality or Reasonableness have their proper scope and limits, and the traditional balance of Reason is reestablished (ibid., p.173).

Toulmin argues that in a situation in which no single mode of theoretical reasoning has hegemony over all others, theories operate simply as ‘models’ of the phenomena they represent (ibid.). Kellert (op. cit., p.85-7) highlights a key implication of chaos theory as the need to adopt a semantic view of theories, which focuses on the construction of models of dynamical systems and theoretical hypotheses expressing a relationship between a model and

a real-world system. Thus, scientific understanding is provided through the search for order in patterns or regularities, focusing on emergent, holistic properties and historical development of systems over time. This is consistent with Reschers' notion of a more modest, empirically-oriented 'philosophical standardism'. Kellert emphasises the contrast with the three traditional main philosophical accounts of scientific understanding: with the 'epistemic' account, in that it does not provide predictions of quantitative detail; with the 'ontic' account, in that it does not seek hidden causal processes; and with the 'modal' account, in that it does not seek to show that events are necessary due to laws of nature (ibid., p.96-111). He concludes "...chaos theory provides us with understanding that is holistic, historical and qualitative, eschewing deductive systems and causal mechanisms and laws" (ibid., p.114).

Complexity and policy action

Elliot and Kiel (1997, p.65) highlight an important implication deriving from the study of chaos, complexity and dynamic systems:

While non-linear dynamics offers new tools and concepts for gaining a greater understanding of social phenomena and policy problem..., it also suggests limitations on the ability of policy analysts and policy makers to change the dynamics of social phenomena.

The literature on complexity consistently emphasises limitations on our ability to predict, plan and control the behaviour of social systems. This implies a degree of modesty in our expectations about 'solving social problems' through our policy efforts because internal system complexity tends to dampen the effect of policy interventions. Elliot and Kiel (ibid., p.73) support the notion of 'gentle policy action' and this is consistent with Rescher's call for modesty and caution: "The answer to the question of how to conduct life in a complex world is – very carefully. Caution, and whenever possible *pre*-caution, should be the order of the day (op. cit., p.195). Rescher supports the notion of 'bounded rationality' and a trial-and-error approach to policy making:

The fact is that in situations of unmanageable complexity, practice in matters of public policy is often guided more effectively by localized experimental trial-and-error than by the theorizing resources of an intellectual technology unable to cope with the intricacy of interaction feedbacks and unpredictable effects (ibid., p.189).

Kauffman (1995, p.253-66) emphasises the need for localised action in seeking to change complex systems in his notion of 'patching'; a system is broken into patches and localised experimentation with solutions may have a better chance of producing an overall optimal change. He suggests that allowing local autonomy in this way promotes learning and may be a 'fundamental mechanism' for achieving adaptive change.

Recent work on processes of societal governance has accommodated the implications of self-regulation and self-steering and the indeterminacy of outcomes (Baumgartner, 1986). Dunsire (1993, p.22) has questioned "...whether social change at an ever-increasing rate, towards ever-increasing complexity, diversity and criticality, can be accommodated by the arrangements societies now make to keep disturbances within bounds and to steer change away from undesired and towards desired directions..." According to Kooiman (1993, p.249-

51), "...traditional ways and means of governance are no longer adequate ... traditional structures of authority and of problem-solving processes have failed." Mayntz (1993) argues that 'governing failure' is no longer seen as due primarily to 'insufficient intelligence' or 'choice of wrong instruments' but rather as attributable to the 'internal dynamics of societal subsystems', which renders them 'cognitively impenetrable' and resistant to 'purposeful influence' (ibid., p.14).

In this context, Mayntz argues, there is a need for greater emphasis on enhancing "...the independent adaptive, reactive and problem-solving capacities of societal actors..." (ibid.). Thus, "...the state needs to be involved in contextual control through the setting of legal norms that can secure the success of collective societal problem-solving by self-regulation and horizontal co-ordination" (op. cit., p.15). The key to effective problem-solving is 'societal self-organization' and, as Kooiman (1993) concurs, the key task of 'interactive governance' is to influence processes of social interaction, seek to balance social forces and interests and enable social actors and systems so as to facilitate such self-organisation. These arguments have clear affinity with the notions of 'gentle policy action' and localised experimentation discussed above.

The notion of 'interactive governance' is also developed by Amin and Hausner (1997). They argue that the idea of society as a web of interlocking networks of affiliation and interaction, structured around a multiplicity of formal and informal institutions, constitutes "...a powerful metaphor for grasping the problems of social complexity" (ibid., p.10). Networking (or 'relational interaction'), involving both state and non-state governance structures, provides a basis for overcoming the rigidities associated with hierarchy - interactive, deliberative networks with a multiplicity of shared values and responsibilities being more discursive and democratic (ibid., p.14-19; Rhodes, 1997). Strategic guidance – the ability to coordinate, arbitrate and facilitate multiple governance networks – is seen as 'the quintessence' of governing social complexity (ibid., p.18).

In Jessop's (1997) notion of 'heterarchic governance' – involving inter-personal, inter-organisational and inter-sectoral network facilitation – guidance constitutes a generic mechanism. Guidance is '...decentralised, cooperative context-steering which is oriented to producing controlled structural change' (ibid., p.109). Guidance is facilitated by '...reflexive monitoring and dynamic social learning...' (ibid., p.111); it implies a role for the state in setting the ground-rules for governance.

Jessop's emphasis on 'reflexive monitoring and dynamic social learning' echoes Giddens' (op. cit.) argument that in a dynamic and rapidly changing world we rely increasingly on the 'reflexive monitoring of action' to guide us. From this perspective a paradox arises. On the one hand, as Smart (1999, p.71) argues "...reflexivity undermines certainties and promotes ambivalence about modern forms of life." In an increasingly complex world, the partial, contingent and fallible nature of social knowledge becomes more evident and this undermines the scope for 'rational guidance' of human affairs. On the other hand, the increasing importance of 'reflexive monitoring of action' indicates an enhanced role for processes of 'social learning' involving the monitoring and evaluation of action in order to understand its consequences and, on the basis of this understanding, learning how to 'improve' action to achieve desired social ends. Thus, whilst we are more aware of the limits of our knowledge of the social world there is nevertheless a greater need for 'reliable' knowledge of the consequences and outcomes of our attempts to change that world. As Smart (ibid.) argues:

...questions concerning political responsibility and ethical decision-making, the difficulties of adjudicating between the expression and pursuit of self-interest and the promotion and adequate provision of the public domain, as well as the problems encountered in everyday social life of making a choice or taking a stand, have if anything become analytically more significant...

Other commentators have recognised the need for an enhanced capacity for learning as a means of reconciling the implications of increasing social complexity with the requirements of effective public policy intervention. Donald Campbell's notion of the 'experimenting society' was founded upon a commitment to innovation, social reality-testing and learning. Such a society "...would vigorously try out possible solutions to recurrent problems and would make hard-headed, multidimensional evaluations of outcomes, and when the evaluation of one reform showed it to have been ineffective or harmful, would move on to try other alternatives" (Campbell and Russo, 1999: 11). As Oakley (2000:320) states:

The experimenting society would be active, preferring innovation to inaction; it would be an evolutionary, learning society and an honest society, committed to reality testing, to self-criticism and to avoiding self-deception; it would be non-dogmatic, accountable, decentralized and scientific...

Dunsire (1986) has argued that under conditions of uncertainty about the ex ante 'correctness' of policy decisions and about capacities to implement policies as intended, there is a need to strengthen the role of evaluation in providing up-to-date, relevant information on actual performance and to build the capacity to take action to modify policy design and implementation in the light of such information. The implications of Rescher's (1998) analysis of complexity, discussed above, point towards increasing need for monitoring and evaluation of 'how matters work themselves out'. In this situation, a major burden is placed upon policy experimentation and evaluation as key social practices in the processes of 'reflexive social learning' whereby we can potentially achieve 'rational guidance' of social change towards collectively desired ends.

Complexity and evaluation

But what kind of evaluation do we need to enhance our capacity for such 'rational guidance' under conditions of high complexity, uncertainty and ambiguity? Dryzek (1990) argues that the predominant response to the challenge of complexity tends to reflect the preoccupations of instrumental rationality, emphasising more sophisticated forms of analysis and technical aides to decision making, in order to maintain control over a problematical environment. This is consistent with Schwandt's (1997) conception of the 'dominant logic of evaluation' as indexed to the 'modernist paradigm of reason', harnessing objective analysis to rational action whereby "...policymakers seek to manage economic and social affairs 'rationally' in an apolitical, scientized manner such that social policy is more or less an exercise in social technology" (ibid., p.74). In this context, evaluation as a modernist project promises "...to tame the unruly world, to bring order to our way of thinking about what does and does not work for improving social life" (Schwandt, 2000a).

On the face of it a focus on assessing 'what works' in particular circumstances is consistent with Rescher's notion of trial-and-error experimentation and Campbell's experimenting society. In the face of complexity, possible policy solutions to social problems are

conjectures. We cannot be sure a priori that they will work, so we tread carefully, evaluate the effects and learn from the results, moving forward in an incremental way. This is Elliot and Kiel's (op. cit.) strategy of 'gentle policy action' or Jessop's (op. cit.) strategy of 'reflexive monitoring and dynamic social learning'. This may be 'bounded rationality' but it involves continued faith in the capacity of evaluation to deliver reliable knowledge of what works.

Indeed, with the flush of enthusiasm in the UK for evidence-based policy making comes a strong emphasis not just on identifying what works (since our Prime Minister declared that that is what counts) but also on understanding why. In effect, a key requirement of evaluation is to produce reliable causal knowledge as the basis for learning – understanding how and why policies work or fail to work so that success can be replicated elsewhere in different circumstances. A recent report by the UK's National Audit Office (2001, p.14) emphasised the role of evaluation in evidence-based policy making, arguing that it can "...help departments learn lessons and share good practice..." and "...identify ways in which the policy can be improved or developed to increase its impact." It also called for improvements in evaluation to make it 'more practical', "...showing what worked well in improving public services and why, and considering what further practical steps were needed to enhance service delivery and improve effectiveness" (ibid.).

There is an assumption here, then, that evaluation studies can deliver quite precise and detailed causal knowledge of how and why interventions work in particular circumstances which will provide the basis for improvements and for the transfer of successful interventions to other settings. Theory-based evaluation (TBE) has attracted considerable attention due to its promise to deliver such knowledge. Indeed, one of the key catalysts of the rise of TBE, Pawson and Tilley's (1997) 'realistic evaluation', has been influential in articulating the nature of the promise, with its emphasis on identifying causal mechanisms in the quest to understand 'what works, for whom, in what circumstances, and why'.

There is not scope here for a detailed discussion of the claims of TBE but I have argued elsewhere (Sanderson, 2002) that these claims need to be treated with some caution in the context of complex policy interventions. In relatively simple interventions, which address a clearly-define problem for a clearly-defined client group through a single mechanism, there is a reasonable prospect of deriving robust causal knowledge. However, much of the current interest in governmental policy circles is focused on addressing complex problems of poverty and social exclusion through multi-instrument, cross-sectoral policy initiatives and it is in just such areas that the current evidence base is weakest and where the highest expectations are held out for TBE. The prospects for developing reliable, generalisable causal knowledge in such areas must be very uncertain.

For example, in relation to urban regeneration, Harrison (2000) points to difficulties in isolating impacts and identifying 'what works' due to the complexity and cross-sectoral nature of interventions and also due to the difficulties of applying "...methodologies and analytical tools that are up to the task of providing reliable and valid evidence" (ibid., p.224). In relation to complex, cross-cutting neighbourhood initiatives, Granger (1998) argues that it is 'virtually impossible' to achieve precise and clear-cut causal attribution. More generally, even Carol Weiss, one of the foremost proponents of TBE, has argued that "...theories-of-change evaluation may be too difficult and time-consuming for any but ... inquiry into one intervening mechanism between inputs and outcomes" (Weiss, 1997). Elsewhere, she has

conclude pessimistically that "...any hope of deriving generalizable findings is romantic" (Weiss, 2000, p.44).

Therefore when faced with the task of understanding how policy interventions in complex social systems work, we may need to adopt rather modest expectations in relation to the standards of causal inference that can be achieved. Indeed, as discussed in the previous section, it can be argued that we need to revise our notion of 'theory' in the context of complex systems, forsaking the 'ontic' conception of explanation through the discovery of causal processes and mechanisms in favour of the 'semantic' view of theories, which focuses on the development of models (cf. Kellert, 1993). On this view, it is futile to seek to identify and trace individual causal processes and mechanisms in complex systems; emergent interaction effects defy such reductionist analysis. Rather, we must accept that analysis will focus largely on the empirical domain, seeking out patterns and regularities. This is not necessarily to deny a realist ontology; rather it is to suggest that in complex social systems it may simply be impossible to penetrate far below the surface layers of 'empirical reality' to access and isolate underlying generative mechanisms.

This position might be seen to underpin Schmid's (1996) conclusions on the application of TBE in the labour market field. Thus, he argues that:

...the composition of the picture has to be improved by more explicit theory-building and theory-testing, although the current state of the art means the complexity of the subject requires some deliberate eclecticism and modesty in terms of general theory; rather than aiming at an axiomatic system of nomological statements (i.e. statements of universal laws), 'pattern recognition' and 'pattern prediction' is the realistic goal... (ibid., p. 205).

So, one important implication of work on complexity is the need to review critically the notion of 'theory' embodied in TBE and, more specifically, our quest for causal knowledge in seeking to answer the question: 'what works, for whom, in what circumstances, and why?' However, it is possible to extend the critique to question the relevance and validity of prevailing notions of 'rationality' to the task of developing our capacity for beneficial policy action in complex social systems. As indicated above, Dryzek (1990) and Schwandt (1997) (and also Fischer, 1990) argue that our thinking about policy making and evaluation is conditioned by an 'instrumental rationality' that focuses on improving the application of objective scientific knowledge and expertise in the selection of means to given ends. Schwandt (2000b) argues that the dominance of 'technical or cognitive-instrumental rationality' has resulted in the 'deformation of praxis' and maintains that social scientific knowledge (which is general and theoretical) cannot provide the primary basis for evaluation judgements under conditions of 'plurality, uncertainty and difference'.

A similar argument is developed by John Forester (1993) in his analysis of planning practice. Instrumental, 'means-ends' rationality leads us to conceive planning practice in a limited and distorted way as 'context-free technical problem solving' in which the focus is on reducing *uncertainty* through scientific evidence and expertise. Forester argues that planning practice should be seen as a communicative process that is inherently political and shaped by rules, conventions and power structures. Not only does such practice draw on sources of knowledge other than explicit scientific evidence (in particular on tacit and experiential practitioner knowledge) but it also necessarily addresses normative considerations. From this perspective,

planning practice is conceived a 'context-dependent practical action' in which the focus is on dealing with *ambiguity* through communicative processes of dialogue.

Now it can be argued that an increasing appreciation of the principles of complexity theory have added force to the relevance of this perspective which calls for an extension in our conception of 'rationality' in two directions. First, it calls for greater pluralism in the cognitive dimension, acknowledging that scientific knowledge and expertise is not the sole basis for rational choice. Second, it calls for the extension of rationality to cover consideration of values, ends and ethical-moral choices. What we are seeking, it is argued, is a 'practical rationality', which can guide us towards appropriate action in complex and ambiguous social contexts. 'Appropriateness', from this perspective, replaces the conventional 'technical' concept of effectiveness; it embodies a notion of balanced judgement taking into account a plurality of forms of knowledge and relevant ethical-moral concerns.

Indeed, it seems to me that the normative dimension of this argument has particular force in contemporary society. Thus, in his analysis of 'late modernity', Giddens' (1990) argues that social activities and relations are being abstracted from their traditional ties to specific contexts, freed from the restraints of tradition and 'local habits and practices'. With the decline of tradition, reflexivity plays an increasing role as the basis for understanding and acting. However, the conditions of 'wholesale reflexivity' undermine traditional 'certainties' and increase uncertainty and ambivalence. It is the notion of ambivalence that is emphasised by Smart (1999) and that chimes with Forester's notion of ambiguity; Smart (*ibid.*, p.71) refers to the 'ambivalent society', which is "...ambivalent about the direction in which it is heading and/or the goals to be pursued... (and) ...the appropriateness and effectiveness of the means employed..." If we acknowledge this normative dimension of complexity, I would argue that we need a broad conception of practical rationality to guide us in coping with it.

Such a view is consistent with Toulmin's (2001) call for a restoration of the 'balance of reason', acknowledging the validity and role of 'practical wisdom' in assessing what is 'reasonable' or 'appropriate' (rather than 'rational') in dealing with human and social problems. Toulmin (*ibid.*, p.2) argues that in recent years there has been "...a revived interest in questions about values that for a while had come to appear foreign even to philosophy":

This turning of the tide points to a future in which the rational demands of scientific technique will be balanced by attention to the demands of human situations in which intellectual or practical skills can reasonably be put to use (*ibid.*)

For Toulmin, a key element in the restoration of the balance of reason is a recognition of the importance of forms of tacit knowledge grounded in skilled and experienced practice, encapsulated in Aristotle's notion of *phronesis*. Such practical wisdom, applied in the form of 'practical-moral reasoning' that Schwandt (1997, p.79) terms 'critical intelligence', can help us, Toulmin (*op. cit.*, p.123) argues, "...in untying the knots in which our lives enmesh us..." with intelligent analysis "...guided by the ideals that make rational assessments stepping stones to reasonable decisions" (*ibid.*, p.213).

If we accept this position and acknowledge the need to nurture the application of practical wisdom in our efforts to improve the world, then we need to broaden the focus of evaluation beyond the 'technical' concerns of measuring effects, identifying causes and assessing 'what works' so that it can help in the practical task of identifying what can be assessed as

‘appropriate’ or ‘reasonable’ courses of action. To do this, it must be acknowledged that the ethical and moral implications of policies and the values and ‘goods’ (and ‘bads’) that they promote are amenable to ‘rational’ consideration and debate (cf. Julnes, et. al., 1998). Broadening the focus of evaluation in this way also involves broadening its methodologies beyond ‘analytical techniques’ to include methods and accompanying institutional frameworks to promote full, free and open normative debate amongst all those with a stake in the policies concerned.

Complexity and evaluation practice

In this final section I shall give very brief consideration to a number of aspects of evaluation practice and the implications that derive from work on complexity. First, to revisit the earlier discussion on theory-based evaluation (TBE), the implication is certainly not that TBE is not a valid and important approach to evaluation in certain circumstances, but rather that we need to be modest in our expectations of what it can deliver in terms of explanation of ‘what works, for whom, in what circumstances, and why’ in the context of complex social systems. This modesty has two dimensions: first, in terms of where and when TBE is applied; and, second, in terms of the type of ‘explanatory knowledge’ that it can deliver. The second aspect was discussed earlier so I won’t repeat it here. As regards the first dimension, I think the point is made by Michael Scriven (1998) who argues for a ‘minimalist’ approach to TBE or, in the words of the title of his paper, “...the least theory that practice requires...” He suggests that theory is usually just not necessary to establish what works and why and that a basic programme logic will often suffice. Moreover, he argues that the primary purpose of evaluation is to establish how well an intervention works; understanding why is ‘a secondary task’ and not always required.

An implication of this argument is that we should be more cautious in our use of the term ‘theory’ in evaluations that seek to understand how and why interventions work or not. Given the conventional association of theory with causal or law-based explanation, a possible implication of the discussion in this paper might be to forego the term in favour of ‘programme model’ or ‘programme logic’. This might signal more realistically the nature of the explanatory knowledge that can be derived in relation to complex interventions.

Nevertheless, where we are working with programme theory/logic, complexity theory has some important implications for how we specify the key assumptions about how the intervention will work to achieve its intended effects in complex social systems. The nature of such specifications clearly influences evaluation design and methodologies; there may be major implications for how key variables are defined and approaches to measurement and analysis. Advocates of the principles of complexity would argue that in many circumstances specifications of programme theory/logic informed by such principles provide a better basis for capturing empirically and analysing processes of social change.

In his application of complexity theory to economics, Ormerod (1998, p.21) provides the example of industrial location. Thus, conventional theory seeks to identify factors which influence rational, profit-maximising firms to locate in particular locations, for example, proximity to suppliers and markets, accessibility and transport costs, costs of land and premises. Moreover, conventional analysis often seeks to isolate the effects of such factors individually, ‘controlling out’ the effects of others. The relationships thus derived (usually assumed to be continuous, well-behaved functions) are extremely unlikely to hold in actual

situations where such factors interact in a dynamic way with outcomes dependent upon particular conjunctions of conditions. Such conventional analysis cannot adequately explain the sudden, rapid growth of industrial conglomeration (as experienced, for example, in Silicon Valley) and policies based upon such analysis are unlikely to be successful. This can explain why industrial sites established in supposedly 'optimal' locations with reference to conventional analysis of location factors may be unsuccessful.

Roberts and Bradley's (2002) work on the evaluation of 'new public management' (NPM) reforms illustrates some implications of a shift in theoretical perspective. They emphasise the importance of what they call the 'problem formulation' stage "...in defining and setting up all that follows..." (ibid., p.21) and are critical of 'discipline-based' research on NPM:

Social systems are highly interconnected open systems in which system behavior is a combination of intentional human action and system processes of self-organisation. Small changes in one element can produce a cascade of consequences, many unintended and unanticipated, that ultimately affect the structure and performance of the whole system... Research on particular aspects of the system puts the emphasis on the parts, not on the whole (ibid.)

They are that in order to analyse fully the effects of NPM reforms, and 'interdisciplinary system-based approach' is required, which addresses the organisation and dynamics of the reform system as a whole, recognising that public institutions and organisations are highly complex multi-level systems linked by a complicated web of interactions to their environment. This perspective leads to the advocacy of network analysis as "...the research methodology, par excellence, that focuses on the organisation of social relations..." (ibid., p.23) and the adoption of the principles of system dynamics to analyse how processes change over time.

The implications of adopting a dynamic systems perspective are considered by Stroup (1997) who highlights a number of key characteristics: embeddedness of systems in their environment or context; time-dependence of system states determined by the interaction of a large number of variables; non-linearity of relationships between variables, such that effects are not proportional to their causes; and sensitive dependence on initial conditions, such that a system's long-term response to a change initiative is heavily influenced by the particular circumstances prevailing when the change is introduced. This latter characteristic in particular causes problems for analysis in that "...sensitivity to initial conditions can easily invalidate results, since the slightest measurement error typically leads to wide variability and poor fitting of observed data to a predicted model" (ibid., p.125).

Acknowledging such characteristics of dynamic social systems produces some important implications for research and evaluation methodologies. Stroup (ibid., p.133) argues that traditional notions of 'independent' and 'dependent' (or 'exogenous' and 'endogenous') variables are undermined by the interdependence between variables in systems and their environments. In the context of social systems in neighbourhoods, Buck (2001, p.2256) argues: "The problem is in effect one of simultaneity. People are influenced by their context and, at the same time, influence that context." Stroup also discusses the influence of non-linearity in producing low levels of explained variance from traditional linear models; from a dynamic systems perspective, the problem becomes one of model mis-specification which no amount of refinement of measurement can address.

Eoyang and Berkas (1998) consider the implications of complexity for evaluation methods in an organisational context, arguing:

Many standard evaluation tools, techniques and methods rely on basic assumptions about linear organizational dynamics (predictability, low dimensionality, system closure, stability and equilibration). Some of these assumptions are not valid when a system enters the regime of a complex adaptive system (CAS) (ibid., p.1).

A key methodological implication is the need to adopt longitudinal designs in evaluation which track the dynamics of system change over time. The approach to measurement must be conditioned by the theoretical specification of the system in question in terms of what is measured, with what frequency and over what timescale to maximise the potential to capture change and understand its relationship to the policy intervention. However, Eoyang and Berkas (ibid., p.12-13) argue that the implications of complexity push us towards measurement over as long a time period as possible and at as frequent intervals as possible, although this must take into account the burden placed upon research subjects (cf. Roberts and Bradley, op. cit., p.29).

Moreover, as indicated above, recognition of the complexity of processes of dynamic system change imply a strong emphasis on capturing the overall pattern of change in the whole system – providing a sound basis of empirical description. Careful attention needs to be given to definition of system boundaries, recognising the potentially high degree of interaction between the system of interest and its ‘environment’ and the fluidity of these definitions. Thus, variables that are initially defined as ‘contextual’ may be affected over time by an intervention. This in turn highlights the problem of unanticipated effects; there is a danger of prior theoretical specifications excluding important effects so it is important that as many aspects of the system and its context as feasible are tracked over time. A further aspect of system comprehensiveness is a multi-level approach, recognising that processes of change operate interdependently at individual, group and whole system levels and that emergent phenomena cannot be reduced to analysis at a ‘lower’ level.

A problem that evaluators face in addressing these characteristics of complex systems is designing an appropriate mix of what are challenging methods and techniques. It is clear that the implications of complexity lead us towards multi-method designs with very heavy demands for high quality data; they require the abandonment of traditional divisions between methodological camps, recognising the need to craft evaluation designs that combine creatively quantitative and qualitative methods. Eoyang and Berkas (op. cit., p.13) are optimistic:

All of these tools and techniques ... provide ways for the evaluator to capture and interpret information about the performance of a human CAS. When these approaches are used in conjunction with more traditional quantitative and qualitative evaluation methods, it will be possible to generate an assessment of a CAS that matches the variety and richness of the system itself.

Perhaps the key implication of complexity, then, is the increased focus on *evaluation design* rather than on methodology per se. Where we are faced with the task of evaluating policy interventions in complex dynamic social systems the key tasks lie in the conceptual specification of the system and its context, the specification of the programme logic/theory, and the crafting of an evaluation design that ‘fits’ this specification and is capable of

capturing and analysing change in the system and how such change is due to the policy intervention.

Given the strong time-dependence of change processes, Eoyang and Berkas (ibid., p.13) argue that it is important that the evaluation design matches the development stage of the system. This emphasises the importance of ‘evaluability assessment’ to determine not just whether the system is at a stage where evaluation is worthwhile, but more specifically to assess the state of the system in relation to a hypothesised path of evolution and determine what approach to evaluation is appropriate, in terms of purpose and design. Moreover, Eoyang and Berkas (ibid., p.7) emphasise the importance of an evolutionary approach to evaluation design, whereby the design is kept under review and modified to adapt to the changes in the system being evaluated. Thus, the relative emphasis on different purposes may shift over time (say, from formative in the early stages, to explanatory in later stages, with a summative emphasis also increasing in later stages) and the design will need to adapt accordingly. Of course, this process of design adaptation calls for careful planning and management of evaluation.

A further important implication of complexity for evaluation design is the need to involve all key stakeholders in the processes of initial design and adaptation. This increases the capacity of evaluators to produce and sustain an appropriate design through learning from key participants in different ‘locations’ in the system and therefore with different perspective on what is relevant and important. It also increases the prospects for learning from the evaluation on the part of such participants and stakeholders, enhancing the role of evaluation as “...an instrument of transformative change...” (ibid., p.14). Indeed, this shift in the role of evaluation to become a key element in learning process is a major implication highlighted by Eoyang and Berkas (op. cit.) and takes us full circle to our earlier discussion of the need for ‘dynamic social learning’ as a fundamental process for societal guidance under conditions of complexity.

References

- Amin, A. and Hausner, J. (1997) *Beyond Market and Hierarchy: Interactive Governance and Social Complexity*. Cheltenham: Edward Elgar.
- Baumgartner, T. (1986) Actors, Models and Limits to Societal Self-Steering, in F. Geyer and J. van der Zouwen, (Eds.) *Sociocybernetic Paradoxes: Observation, Control and Evolution of Self-Steering Systems*. London: Sage Publications.
- Buck, N. (2001) Identifying Neighbourhood Effects on Social Exclusion. *Urban Studies*, Vol. 38, No. 12, pp. 2251-2275.
- Campbell, D. T. and Russo, M. J. (1999) *Social Experimentation*. Sage Classics 1. Thousand Oaks: Sage Publications.
- Dryzek, J., S. (1990) *Discursive Democracy: Politics, Policy and Political Science*. Cambridge: Cambridge University Press.

- Dunsire, A. (1986) A Cybernetic View of Guidance, Control and Evaluation in the Public Sector, in F-X. Kaufmann, G. Majone, V. Ostrom and W. Wirth (Eds.) *Guidance, Control and Evaluation in the Public Sector*, Berlin: Walter de Gruyter.
- Dunsire, A. (1993) Modes of Governance, in J. Kooiman (Ed.) *Modern Governance: New Government – Society Interactions*. London: Sage Publications.
- Dupre, J. (2001) *Human Nature and the Limits of Science*. Oxford: Clarendon Press.
- Elliot, E. and Kiel, L. D. (1996) Introduction, in Kiel, L. D. and Elliot, E. (Eds.) *Chaos Theory in the Social Sciences: Foundations and Applications*. Ann Arbor: University of Michigan Press.
- Eoyang, G. and Berkas, T. (1998) Evaluating Performance in a Complex Adaptive System, in M. Lissack and H. Gunz (eds.) *Managing Complexity in Organizations*. Westport, Connecticut: Quorum Books.
- Fischer, F. (1990) *Technocracy and the Politics of Expertise*. Newbury: Sage Publications.
- Forester, J. (1993) *Critical Theory, Public Policy and Planning Practice: Toward Critical Pragmatism*. New York: State University of New York Press.
- Giddens, A. (1990) *The Consequences of Modernity*. Cambridge: Polity Press.
- Granger, R. C. (1998) Establishing Causality in Evaluations of Comprehensive Community Initiatives, in K. Fulbright-Anderson, A. C. Kubisch and J. P. Connell (eds.) *New Approaches to Evaluating Community Initiatives, Volume 2: Theory, Measurement and Analysis*. Washington D. C.: Aspen Institute.
- Harrison, T. (2000) Urban Policy: Addressing the Wicked Problems, in H. T. O. Davies, S. M. Nutley, and P. C. Smith (eds.) (2000) *What Works? Evidence-Based Policy and Practice in Public Services*. Bristol: Policy Press.
- Harvey, D. L. and Reed, M. (1996) Social Science as the Study of Complex Systems, in Kiel, L. D. and Elliot, E. (Eds.) *Chaos Theory in the Social Sciences: Foundations and Applications*. Ann Arbor: University of Michigan Press.
- Jessop, B. (1997) The Governance of Complexity and the Complexity of Governance: Preliminary Remarks on Some Problems and Limits of Economic Guidance, in A. Amin and J. Hausner (1997) *Beyond Market and Hierarchy: Interactive Governance and Social Complexity*. Cheltenham: Edward Elgar
- Julnes, G., Mark, M. M. and Henry, G. T. (1998) Promoting Realism in Evaluation: Realistic Evaluation and the Broader Context, *Evaluation*, Vol. 4, No. 4, pp.483-504.
- Kauffman, S. (1995) *At Home in the Universe: The Search for Laws of Complexity*. London: Penguin Books.
- Kellert, S. H. (1993) *In the Wake of Chaos: Unpredictable Order in Dynamical Systems*. Chicago: University of Chicago Press.

- Kooiman, J. (1993) Findings, Speculations and Recommendations, in J. Kooiman (Ed.) *Modern Governance: New Government – Society Interactions*. London: Sage Publications.
- Majone, G. (1989) *Evidence, Argument and Persuasion in the Policy Process*. New Haven: Yale University Press.
- Mayntz, R. (1993) Governing Failures and the Problem of Governability: Some Comments on a Theoretical Paradigm, in J. Kooiman (Ed.) *Modern Governance: New Government – Society Interactions*. London: Sage Publications.
- Marion, R. (1999) *The Edge of Organization: Chaos and Complexity Theories of Formal Social Systems*. Thousand Oaks: Sage.
- National Audit Office (2001) *Modern Policy-Making: Ensuring Policies Deliver Value for Money*. London: NAO.
- Nicolis, G. and Prigogine, I. (1989) *Exploring Complexity: An Introduction*. New York: W. H. Freeman and Co.
- Oakley, A. (2000) *Experiments in Knowing: Gender and Method in the Social Sciences*. Cambridge: Polity Press.
- Ormerod, P. (1998) *Butterfly Economics: A New general Theory of Social and Economic Behaviour*. London: Faber and Faber.
- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, London: Sage Publications.
- Prigogine, I. (1996) *The End of Certainty: Time, Chaos, and the New Laws of Nature*. New York: The Free Press.
- Reed, M. and Harvey, D. L. (1992) The New Science and the Old: Complexity and Realism in the Social Sciences. *Journal for the theory of Social Behaviour*, Vol. 22, No. 4, pp. 353-380.
- Rescher, N. (1998) *Complexity: A Philosophical Overview*. New Brunswick: Transaction Publishers.
- Rhodes, R. A. W. (1997) *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability*. Buckingham: Open University Press.
- Roberts, N. C. and Bradley, R. T. (2002) Research Methodology for New Public Management, *International Public Management Journal*, Vol. 5, No. 1, pp.17-51.
- Sanderson, I. (2002) Does it Work? Is it Appropriate? Instrumental and Practical Rationality in Evidence-Based Policy Making. Paper presented at European Evaluation Society Fifth International Conference, *Three Movements in Contemporary Evaluation: Learning, Theory and Evidence*, Seville, Spain, 10-12 October, 2002.

Schmid, G. (1996) Process Evaluation: Policy Formation and Implementation, in G. Schmid, J. O'Reilly and K. Schomann (Eds.) *International Handbook of Labour Market Policy and Evaluation*. Cheltenham: Edward Elgar.

Schwandt, T. A. (1997) Evaluation as Practical Hermeneutics. *Evaluation*, Vol. 3, No. 1, pp. 69-83.

Schwandt, T. A. (2000a) Meta-Analysis and Everyday Life: The Good, The Bad and The Ugly, *American Journal of Evaluation*, Vol. 21, No. 2, p.213.

Schwandt, T. A. (2000b) Further Diagnostic Thoughts on What Ails Evaluation Practice, *American Journal of Evaluation*, Vol. 21, No. 2, p.225.

Scriven, M. (1998) Minimalist Theory: The Least Theory That Practice Requires, *American Journal of Evaluation*, Vol. 19, No. 1.

Smart, B. (1999) *Facing Modernity: Ambivalence, Reflexivity and Morality*. London: Sage Publications.

Stroup, W. F. (1997) Webs of Chaos: Implications for Research Design, in R. A. Eve, S. Horsfall and M. E. Lee (eds.) *Chaos, Complexity and Sociology: Myths, Models and Theories*. Thousand Oaks: Sage Publications.

Toulmin, S. (2001) *Return to Reason*. Cambridge, Ma.: Harvard University Press.

Weiss, C. H. (1997) How can Theory-Based Evaluation Make Greater Headway? *Evaluation Review*, Vol. 21, No. 4, p.501-25.

Weiss, C. H. (2000) Which Links in Which Theories Shall We Evaluate? in P. J. Rogers, A. Petrosino, T. A. Heubner and T. A. Hasci (Eds.) *Program Theory Evaluation: Practice, Promise, and Problems*, *New Directions in Evaluation*, No. 87, pp.35-45.