

This document is the Accepted Manuscript version of the following article:
Chris Mowles, 'Complex, but not quite complex enough: The turn to the
complexity sciences in evaluation scholarship'.

The final, definitive version of this paper has been published in *Evaluation*, Vol.
20 (2): 160-175, April 2014, doi: <https://doi.org/10.1177/1356389014527885> ,
published by SAGE Publishing. All rights reserved.

Complex, but not quite complex enough: the turn to the complexity sciences in evaluation scholarship.

Abstract

In recent years there has been a turn to the complexity sciences by evaluation scholars as way of producing richer theories of the social. However, and at the same time, what might be considered some of the more radical insights from the non-linear sciences are sometimes lost when they are represented within what I understand as a dominant paradigm; that evaluation is primarily a rational, detached activity based on design. This is particularly the case with evaluation scholars who adopt a contingency approach to complexity, and who argue that social interventions are not always complex, or may only be partly so. In critiquing the way that the complexity sciences are adduced I am making stronger claims for their significance, but also urging greater caution. Lack of precision in the way theories are taken up from the complexity sciences in evaluative terms may lead both to under- and over-claiming their import. In under-claiming scholars may put forward the idea, for example, that emergence is a special category of human activity and means the opposite of being planned. In over-claiming, they may argue that 'complexity science' supports any number of arguments which seem unrelated to any particular manifestation of the non-linear sciences. Insights may also get lost an appeal to abstract systems, mechanisms and 'levels' of reality. In this article I claim that human interaction is always complex, that emergence can be understood as the interweaving of intentions, and that evaluation is a social activity like any other, and is therefore not exempt from the evaluative process. Making claims for the replicability of 'successful' social interventions is a probabilistic undertaking if we accept one of the central insights from complex adaptive systems theory, that global patterns may tell us very little about the micro-processes that brought them about.

Key words: complexity sciences, complex adaptive systems theory, social theory, systems theory.

Article

During the last 10-15 years there have been repeated appeals to the complexity sciences to inform evaluative practice in the scholarly literature. This partly reflects the increased ambition of many social development and health programmes which are configured with multiple objectives and outcomes (Stern et al, 2011; Forss et al, 2011, Connell et al, 1997) and the perceived inadequacy of linear approaches to evaluating them. It could also be understood as a further evolution of the methods vs. theories debate (Chen and Rossi, 1989; Weiss, 1995, 1997, Pawson and Tilley, 1997; Stame, 2004) which has led to theory-based approaches becoming much more widely taken up in the evaluative practice. Although discussion over paradigms have clearly not gone away (Banerjee and Duflo, 2012; Hedström, 2005) the turn to the complexity sciences as a resource domain for evaluative insight could be seen as another development in producing richer theories better to understand, and make judgements about, complex reality. However, some evaluators are understandably nervous about the challenge of what they perceive as being the more radical implications of assuming that non-linear interactions in social life may be the norm, rather than the exception.

This article calls into question how far a turn to the complexity sciences in the evaluation literature marks a departure from more conventional understandings of the role and function of evaluators and to evaluation scholarship, and asks whether some of the more radical implications have simply been subsumed into the more orthodox ways of thinking by what Fleck (1979) referred to as the prevailing 'thought collective'. Fleck, an intellectual forebear of Thomas Kuhn (1996), described the ways in which groups of scholars, committed to understanding the world in a particular way, resist the rise of new ideas by either ignoring them or rearticulating them in terms of the prevailing orthodoxy. The article is principally theoretical and sets out to enquire into how the complexity sciences are understood by a variety of evaluation scholars, both in relation to the original scientific discipline from which the theories originate, and in terms of how they are then understood in social terms. Further, the article goes on to ask what difference appealing to the complexity sciences makes to the prescriptions that scholars recommend for evaluative practice. Lastly the article explores what might be considered some radical alternatives to evaluation and complexity scholarship to date, which, in my view is still rather attenuated, apart from a few exceptions.

My argument will focus on scholarship which makes a direct appeal to the complexity sciences, rather than using the term as, for example Stame (2004), Spicer and Smith (2008) and van der Knaap (2011) do, in an every day sense. These latter are no less insightful about the difficulty posed for evaluation as a discipline by complex social programmes. For example, in a helpful article taking a historical perspective on the evolution of the evaluation paradigm debate Stame argues that social development initiatives can be complex because they are stratified, with social actors embedded in their own contexts, and because each aspect that could be examined is multi-faceted (2004: 63). Stame thinks of complexity in terms of multi-layered governance with EC policy makers wanting to achieve certain outcomes in general, but being far removed from the contexts in the different European countries where the programmes are implemented. The difficulties that she identifies arising from the paradox of global intent and local implementation are to be found repeated in much of the literature that draws on the complexity sciences directly. Meanwhile, Van der Knaap writes interestingly about the importance of the evaluator paying attention to both the unexpected and the unwanted, in what he terms 'responsive' performance audit and argues that from a variety of different perspectives evaluation orthodoxy is under pressure to be more participative, to be more thoughtful and to take more account of power and politics.

Nonetheless, this article will discuss the extent to which the complexity sciences have come to disturb the discipline of evaluation, and will consider the ways that scholars think that they are, or are not helpful.

Trends in the complexity and evaluation scholarship

I will discuss three characteristics about the complexity and evaluation literature which have struck me because they seem to be at the heart of the struggle over the complexity sciences and their import for evaluation. The first thing to notice is the number of scholars (Rogers, 2008, 2011; Patton, 2011; Forss and Schwartz, 2011; Funnel and Rogers, 2011; Marra, 2011; Ling, 2012; Stern et al., 2012) who suggest more or less strongly that the complexity sciences, however they are taken up, may be a perspective only applicable in particular circumstances and at particular times according to the evaluator's assessment. Programmes which need evaluating, the authors claim, are either simple, complicated or complex, or complex programmes may have simple or complicated parts. It is considered to be a perspective, as Ling points out, that 'can somehow be grafted onto more conventional approaches' (2012: 84) dependent upon circumstances. In a domain which is

replete with a dizzying array of tools, techniques and perspectives all offered with propositional (if, then) logic, a complexity perspective is another weapon in the rational evaluator's armoury.

A second observation is the tendency to homogenise the various manifestations of different natural science complexity disciplines and to talk about them as though they could be understood as one theory with a common set of shared characteristics. Alternatively, the articles proceed as though the complexity sciences have one homogenised theory, but actually talk selectively about some characteristics of particular manifestations of the complexity sciences. While recognising the important move seriously to consider the implications for evaluation practice of non-equilibrium and non-linearity, this article will question whether it is enough just to call on ideas from the complexity sciences in general with the danger that important implications get lost in translation. There is also a danger of over-claiming what 'complexity science' does or does not support in our theorising about the world.

And the third tendency, partly connected to the first and second points above, is a taken for granted functionalism, that assumes that social and/or health development interventions are helpfully construed as systems. Interaction is then understood as taking place between entities, agents, even institutions operating at different 'levels' of the system. The parallel between, say, complex adaptive systems theory, and the intervention understood as system, becomes a neat one and allows scholars to avoid explaining their theory of social action, or to interpret complexity theories from the perspective of social theory and thus to read into them more than they sustain.

Simple, complicated and complex – choosing when to 'apply' complexity

Scholars who have an interest in the complexity sciences have an understandable need to define what they are talking about both for themselves and their readers, and this has no doubt motivated them to draw on what has become known as the Stacey matrix (Stacey, 1992). Stacey's matrix represents a contingency theory of organisations understood as complex adaptive systems and suggests that the nature of the decision facing managers depends on the situation facing them. In situations of great uncertainty and high disagreement, then conventional linear/rational methods of analysis and decision-making are inadequate, Stacey argues. Variations on Stacey's idea of presenting complexity as contingent decision-making have been reproduced by others,

most notably Glouberman and Zimmerman (2002), who seem to have gained purchase amongst evaluation scholars such as Rogers (2008, 2011), Funnel and Rogers (2011) and Patton (2011), which has then led to others citing these ideas.

Glouberman and Zimmerman's proposal is that social problems are of three kinds: simple, complicated and complex. Simple problems require following a recipe, which, once mastered, carries with it a very high assurance of success in future. Complicated problems 'contain subsets of simple problems but are not merely reducible to them. Their complicated nature is often relatedto the scale of a problem like sending a rocket to the moon' (2002:1). Complex problems are ones like raising a child, where there is no formula to follow, and 'success' with one child does not guarantee success with the next.

This is the kind of formulation which may look helpful on first reading but does not stand up to much careful investigation. Nor does it become more credible because it is widely taken up and endlessly repeated. It is hard to conceive of sending a rocket to the moon, except in the very narrow sense of being able to see whether one has landed on the moon or not, as being anything other than a complex undertaking. Inevitably on each occasion it will have involved widespread mutual adaptation and improvisation, disagreements, lacunae, the unexpected and the contingent, and with occasional catastrophic consequences (Apollo 13 and the Challenger disaster). Even following rules like a recipe, to draw on the Canadian philosopher Charles Taylor (1999), is a highly social process where the rules inform practice and practice informs the rule. There is no recipe so clear that it is completely obvious what to do in every situation, and rules are 'islands in the sea of our unformulated practical grasp of the world' (1999: 34).

If, as this article claims, the heuristic does not seem to support the weight of expectation freighted upon it, how might one account for its continued appeal? Although there seems to be some agreement that insights from the complexity sciences help us understand why social activity is unpredictable, to consider evaluation practice, which is also a social activity, in the same light radically decentres it: it can no longer be grounded in the certainties of the rational, designing evaluator. That is, if you accept Smith's (1989) contention that without an explicit model laying out goals and measurable objectives a programme cannot be evaluated, then theories of change methods, all based on propositional logic models, immediately become problematic. Each

of the proponents of Glouberman and Zimmerman's framework acknowledges this to a greater or lesser extent. Rogers (2008) argues that an evaluator needs to make their logic model more 'flexible', which appears to mean developing a series of logic models and being prepared to evaluate what she terms 'emergent' aspects of the programme. Patton (2011), meanwhile, is more or less radical, depending on whether the evaluation is deemed 'developmental' or not, according to the simple, complicated or complex heuristic. Similarly, Marra (2011: 331) concludes that while 'complexity may potentially offer a valuable framework to understand complex systems, a complexity perspective cannot be applicable across all evaluation settings.' In general then, Glouberman and Zimmerman's schema allows evaluators in the mainstream to claim that the complexity sciences may be quite helpful but only in circumstances of their own choosing. They continue to cleave to what John Dewey (2005) referred to as a 'spectator theory of knowledge', which sustains a separation between the observer and the thing observed.

There are scholars who take up the complexity sciences and do problematize the activity of evaluators more consistently than those who adopt the tripartite heuristic, such as Callaghan (2008) and Sanderson (2000, 2009), whose work will be discussed more thoroughly below.

One theory, several theories and over-claiming

Whether it is helpful to enquire into the social with the support of theories derived from the natural sciences of complexity is contested. For example, and as Mowles et al. (2008) point out, in the domain of organisational theory, a number of scholars who are sympathetic to the idea of a complex social reality nonetheless problematise the instrumental application of complexity theory to social life (Chia, 1998; Tsoukas and Hatch, 2001; Tsoukas, 2006). They reject the direct transfer of models from the natural sciences and explore how, if at all, complexity science ideas may be located in the context of established social theory relevant to organisations. Chia, for example, understands the application of complexity theory to social contexts as just another manifestation of scientific reductionism.

If one is going to take up the complexity sciences in social settings, then, it is important to proceed with caution and to be as clear as possible how one is interpreting them. As pointed out above, some scholars develop a generic understanding of what they understand the complexity sciences to be saying, while others

draw on examples of particular theories. Just to take a sample of the sciences of complexity to which different authors refer, Sanderson (2000) and Callaghan (2008) allude to Prigogine's (1977) dissipative structures, but without ever directly referring to him or what he took his work to mean. Meanwhile, Rogers (2008) and Westthorp (2012) refer to complex adaptive systems theory, as does Davies (2004, 2005) until he abandons it in favour of Social Network Theory as a means of drawing aggregate lessons from evaluations of large projects. From a contingency perspective discussed above, Patton (2011) draws extensively on a wide range of scholarship on the complexity sciences, tipping points, simple rules, CAS, without offering a view as to whether one particular branch of the complexity sciences is more helpful for understanding the discipline of evaluation than another.

Does it really matter, or is it just a form of hair-splitting to argue that which theory and the way it is taken up, counts? Clearly it matters to some: Stern (2008) argues that the term complexity is 'frequently and carelessly bandied about these days'. There is a need to understand the implications of the complexity sciences with greater precision. In the development of the disciplines of evaluation, particularly those claiming to be theory-based, it is probably important to know what the theories being adduced actually claim to be revealing about nature, and to be able to make distinctions between one theory and another. To demonstrate what the difference is that makes a difference will take at least a brief explanation.

In applying a heuristic, drawing on the work of Peter Allen (1998a, 1998b), to explain the way that different manifestations of the complexity sciences have evolved over time, Stacey (2011) demonstrates in step-wise fashion the different assumptions involved in a spectrum of mathematical or computer-based models of complexity. What we might term Step 0 represents an equilibrium model commonly found in classical physics and economics which assumes: a) a system with a boundary comprising interacting entities, b) that the entities themselves are homogenous, c) that the interactions between entities are assumed to be occurring at an average rate and d) that the system is assumed to move towards equilibrium.

Step 1 following the development of complexity models in the natural sciences involves taking away assumption d), that the system is moving toward equilibrium, and substitutes non-linear for linear equations. The output of one equation simply feeds into the next iteration of the equation and is the basis for modeling

the movement of mathematical chaos, understood differently from the meaning of the word chaos in every day parlance, over time. For particular parameter values the model produces perfectly stable, predictable movement, known as a point attractor. At other values the model produces predictable patterns of peaks and troughs, known as a period two attractor. Meanwhile at different values again the model demonstrates explosively unstable behaviour. What Stacey identifies as important in models of mathematical chaos is that between parameter values which induce stability and high-dimensional chaos the system moves towards what is known as a strange attractor, which appears to be random but on closer inspection reveals itself to be a pattern which is both regular and irregular at the same time. These are not separate states, regular then irregular, but create a mutually constitutive, paradoxical movement over time of stable instability, or unstable stability. Models of mathematical chaos are very sensitive to small changes in initial conditions where a very difference in the data input can escalate to large scale change in patterning over time. This makes prediction as to how the model will evolve highly problematic.

Step 2 of the evolution of complexity modelling involves dropping two assumptions from the classical paradigm, d) that the system is moving towards equilibrium and c) that interactions are occurring at an average rate. Stacey explains Prigogine's theory of dissipative structures using this step, which models convection patterns in laboratory experiments when heat is applied to certain liquids. At certain critical values the stability of the liquid breaks down and can spontaneously jump to a different state following non-average interaction with the environment, in this case with the application of heat. Unlike the model of mathematical chaos the process of change is not fully self-referencing, but is dependent upon absorbing energy from the environment, without which the change dissolves. But the liquid is also demonstrating a paradoxical ability to create new order from disorder and to self-organise at particular parameter values.

Complex adaptive systems (CAS), agent-based models run on computers, have been developed to model evolutionary behaviour of diverse populations, or to reflect the way that, say, ants organize themselves in an ant colony, or even how the brain works. In other words, they are temporal models which change qualitatively over time and attempt to explain how order emerges from apparent disorder, without any overall blue-print or plan. They model dynamic behaviour. CAS can operate both at Step 2, as above, and at Step 3 of Allen's heuristic, where, in Step 3 the three assumptions d) equilibrium, c) average rate interaction and b)

homogenous entities are all dropped. Both types of CAS models differ from mathematical chaos and dissipative structures because they are not macro models but attempt to describe how global patterns arise from local agent behaviour. The relation between the local and the global is a question, as we have noted before, which preoccupies evaluation scholars too (Stame, 2004; Davies, 2005). A Step 2 CAS, most famously Reynold's 'boids' simulation (1987), models birds flocking. Homogenous boids, bit-strings of computer code, follow three simple instructions and by doing so mimic flocking patterns. These instructions are for each boid to match velocity with other boids in the vicinity, to head towards the densest part of the population of other boids in the vicinity, and to keep equi-distant from their neighbours. With these simple rules a self-organising global flocking pattern emerges as a consequence of boids acting locally.

This latter example has been widely taken up in both organizational and evaluative literature with the recommendation that managers, project leaders, or evaluators should somehow just introduce simple rules to 'encourage emergence to happen'. However, and as Stacey points out, all the boids are the same and they are all following the same rules: as a consequence the pattern of emergence never evolves beyond flocking. It is important to consider how helpful such a proscription is for social life, where all human beings are unique, and even the simplest rules are open to misinterpretation. Stacey argues that it is only when the agents are diverse, their interactions are non-average and non-linear, a step 3 model, that what we might describe as truly evolutionary and novel behaviour emerges. In step 3 the model takes on a life of its own and produces patterns that the programmer could not have predicted.

Ray's (1992) simulation of hosts and predators is an early example of evolutionary step 3 CAS modelling, and the discipline has been taken up in Miller and Page's (2007) work modeling complex social dynamics, Hedström's analytical sociology (2005, 2008) and Allen's models of evolving industries (Allen et al., 2006). What I am terming step 3 CAS comes closest to what Pawson and Tilley (1997), from the domain of evaluation, describe as generative causality. In other words, global patterns arise as a consequence of what each individual agent is doing with other similar agents acting locally, but the global pattern may give no insight into that local interaction. There are significant differences between Pawson and Tilley's idea of generative causality and CAS, however, which I will explore below, but this is partly because, as Hedström expresses it: 'aggregate patterns say very little about the micro-level processes that brought them about' (2005: 99). Just to

restate, emergence is taken to mean a global order which arises from, but is not reducible to, the sum of all local activity which in CAS terms is always governed by rules.

It is evident from Stacey's careful exploration of the different manifestations of the complexity sciences that each development of non-linear, non-equilibrium modelling undermines a more reductionist account of stability and change. However, each has its own characteristics and limitations, and can lead to different conclusions about how we might understand complex reality which is changing over time. Nor can we assume, as Stacey reminds us, that we can simply and unproblematically 'apply' these insights to the social as though agent-based models, mathematical models of weather patterns or chemical reactions, were the same as social activity. Models are helpful in supporting as to think about real world problems, but as CAS practitioners de Marchi and Page (2008) remind us 'mathematical models uncover fundamental truths about mathematical objects and not much about the real world'. It is for this reason that Stacey is most interested in the insights of, in particular, step 3 CAS models and takes them up by analogy to explain the stable instability of social life. He develops his theory of complex responsive processes by drawing on similar insights from the social sciences, in particular the perspectives of the sociologist Norbert Elias (2000) and social behaviourist GH Mead (1934). He turns to CAS, because unlike mathematical chaos and dissipative structures, step 3 models are agent-based and temporal, modelling the changing and paradoxical relationship between the local and the global: local interaction between agents forms, and at the same time is formed by, global patterning and evolves qualitatively over time.

There are numerous examples in the evaluation literature where scholars do not explain fully how they are linking their ideas to particular complexity sciences and argue too loosely for what the complexity sciences can or cannot support theoretically. It allows them either to over- or under-claim their significance for evaluative work. Alternatively, they interpret the complexity sciences to promote the default conception of evaluation as a rational, stable activity which can absorb identifiable pockets of complexity.

As an example of the former, Marra, draws on 'complexity science' to make some quite grand statements about evaluative practice suggesting that it tells us 'that all collections of individuals, including groups, inherently produce a hierarchical structure of roles.' (2011: 326) or that it supports the need to be inclusive and sharing

(op. cit: 327) or that it supports the need for evaluators to distribute information and create 'collective intelligence' (op. cit: 330). Ling (2012) argues that drawing on complexity science, which he claims privileges the whole over the parts, helps reduce uncertainty, as does Spinatsch (2011), as at times in an extremely wide-ranging and sometimes contradictory volume, does Patton (2011). It is difficult to trace how these scholars reached the conclusions they did from the sources they adduce. Meanwhile, as examples of under-claiming to fit a pre-existing set of certainties, those scholars already predisposed to a contingency approach to complexity (Rogers, 2008, 2011; Patton, 2011) suggest complexity is a 'lens' or a framework to be applied if helpful, and take emergence to mean the opposite of being tightly planned. Complex, dynamic and emergent interventions are a particular category of social programme, or even a part of social programme, which need a special and dynamic response, with evaluators trying to feed back data and information in real time (Ling, 2012; Marra, 2011; Patton, 2011). In portraying emergence as a special phenomenon they have implicitly dismissed the idea that the human interaction is always complex, and that emergence, which we might understand in social terms as the interplay of intentions, is always happening, whether a social programme is tightly planned or not.

On the other hand there are also scholars, such as Callaghan (2008), Sanderson (2009) and Westhorp (2012) who argue strongly for drawing on insights from the complexity sciences more generally to inform evaluation practice, rather than understanding the insights to refer only to special cases. Callaghan and Westhorp argue in favour of what they term 'complexity-consistent' approaches, while Sanderson argues that accepting disequilibrium as the norm rather than the exception implies greater modesty in social policy formulation as well as the practice of evaluation. I will investigate their work again below.

Complexity and degrees of abstraction – the limitations of systemic theorising

Evaluation scholars abstract to varying degrees from the social programmes they are invited to evaluate.

Perhaps the highest degree of abstraction is demonstrated by those evaluators using experimental methods who are concerned to draw statistical distinctions between a 'treatment group' and a comparator group which is randomly selected. Experimentalists are generally disinterested in social theory and think of causality in terms of independent and dependent variables. Meanwhile, adherents of Theories of Change (ToCs) made popular by the Aspen Institute (1997), draw on propositional logic and represent social change in the form of entity-based logic models showing the linear development of social interventions towards their conclusions.

Additionally, however, they will often point to the importance of participation and involvement of the target population of programmes to inspire motivation (Funnel and Rogers, 2011). In this sense TOCs are a hybrid of functionalism and emancipatory social theory.

Realist evaluators claim to be interested in 'generative' theories of causality, i.e. ones which open up the 'black box' of what people actually do to make social programmes 'work' or not. Realistic evaluation draws on Bhaskar's critical realism (1978) as taken up and developed by Pawson and Tilley (1997) and Pawson (2006) and is the theory most often linked to the complexity sciences, particularly CAS. For example, Callaghan (2008), Marra (2011), Westhorp (2012), Barnes et al. (2003) all try to reconcile realistic evaluation and CAS. In doing so they adopt a functionalist, systems-based understanding as a default position and argue that interactions between human beings take place as 'mechanisms' and have an effect at different 'levels' of reality. Previously in this article we have made a conceptual link between step 3 CAS and realistic evaluation because they both have an understanding that stability and change does not arise because of 'variables', the staple of experimental methods, nor does it proceed with propositional logic as in ToC, but as a result of what agents are doing in their local interactions with other agents. Step 3 CAS are relational models demonstrating how patterns emerge over time because of ensembles of interacting agents. So from a realistic perspective and in the words of Pawson and Tilley:

Realists do not conceive that programmes 'work', rather it is the action of stakeholders that makes them work, and the causal potential of an initiative takes the form of providing reasons and resources to enable programme participants to change. (1997: 215)

Realistic evaluators argue that interventions do or do not achieve what they set out to because of a combination of context, mechanism and outcomes (CMO). It is concerned with finding what works for whom and in what circumstances and then extrapolating a detailed and evolving explanation to other contexts. In Pawson's words it is predicated on the 'steady accretion of explanation' (2006: 176) about a reality which exists independent of the evaluators who are enquiring into it. Callaghan (2008) adds as further development on the idea of a mechanism, that what people are doing locally in their specific contexts to make social projects work

is to negotiate order, and Westhorp (2012) recommends trying to identify the local 'rules' according to which people are operating as a way of offering richer evaluative explanations of what is going on.

It is easy to see the appeal of the link between a realistic evaluator's interest in what people are doing to make a project work, through negotiated order or rule-following, and CAS. Realistic evaluation has much to recommend it in terms of its insistence on the importance of the particular history and local context of social interventions, and that prediction and questions of validity in different contexts are made highly problematic. However, some of the more arcane aspects of critical realism are in danger of covering over what we might think of as the radical implications of CAS. Rather than opening up the black box of causality realistic evaluators, in Norbert Elias' words (1978: 73), seem to use a mystery to explain a mystery when they draw on the concepts of systems to describe the way that contexts and mechanisms work. For example, Pawson argues that social interventions are 'complex systems thrust amid complex systems' (2006: 168), and that: 'A sequence of events or a pattern of behaviour are explained as being part of a system and the mechanism tells us what it is about that system that generates the uniformity' (2006: 23). According to Westhorp: 'Both complexity theory and a realist philosophy of science understands reality as comprising multiple, nested, open systems...' (2012: 406). For Barnes et al. (2003: 277): 'In Health Action Zone terms we can see the 'context' as being made up of other system levels that interpenetrate...'. In my understanding of CAS models there is nothing to suggest that they are either open, nested, or have multiple levels. The global patterning that emerges may tell us very little about the local interaction that has brought it about: even if we were able to identify local rules, or 'generative mechanisms', they would not necessarily help us, since there may be no obvious connection between local and global 'uniformity'. Introducing functional abstractions, 'system', 'levels', 'mechanisms', covers over as much as it reveals about what may be happening in a social development intervention, and promises more than it can deliver if we are to take the insights from CAS seriously. Rather than being concerned with static, entity-based and spatial representations of complex reality where causal powers are attributed to machine-like mechanisms, CAS models are helpful in understanding qualitative changes in ensembles which change over time.

Of course, realistic evaluators are not the only evaluation scholars to understand what they are doing in systemic terms, no matter how much the idea of a system is problematized: i.e. scholars often claim that

despite using the term, they do not think it is easy to know where the boundary of a system is, or claim that systems are open, or nested, or intersecting with other systems or whatever. It is only a short step to begin thinking that if the idea of a system in social terms is so problematic, then perhaps it would be preferable not to use it at all, but to find some other way of paying attention to, or describing what happens when social development interventions occur. Part of the explanation for the persistence of systemic abstractions may be that they protect the discipline of evaluation by separating the evaluator from the object to be evaluated. In this sense, and despite the encouragement of a variety of evaluation scholars to value reflection, reflexivity and multiple views of reality (Marra, 2011; Patton, 2011; Barnes et al, 2003) this decentering radicalism rarely takes in the discipline of evaluation itself, with some exceptions (Sanderson, 2009; Callaghan, 2008). This is not to argue that evaluators, particularly in the realistic school of evaluation are unaware of the way that they influence social interventions, by learning then 'teaching back' as Pawson and Tilley (1997) express it. To a degree, then, evaluation scholarship takes refuge behind its abstractions and takes what the philosopher Thomas Nagel (1986) described as 'a view from nowhere', by which I understand him to mean that by abstracting away from ourselves as subjective thinkers we leave out precisely what we need to explain. Even those evaluation scholars, who problematize more positivistic perspectives on their discipline only go so far in developing how much these non-linear sciences apply to them and what they are doing in the practice of evaluation.

Taking one more step in linking evaluation with insights from the complexity sciences

In elaborating the perspective of complex responsive processes Stacey (2012, 2011), Stacey, Griffin and Shaw (2002) and Mowles (2011) draw on the complexity sciences to take a more radical position than any of the scholarship cited above, arguing for what we might term step 4. That is to say, Stacey and colleagues argue that the complexity sciences have important implications for the way that we think about social life, but that CAS models only take us so far. CAS models are, after all, merely computer-based simulations of living reality. They argue that in moving from computer modelling to theories of the social, but by preserving some of the insights by analogy, it might be helpful to think of social interaction as tending neither towards equilibrium nor as linear. Social life always takes place local between diverse individuals who have their own history and multiple understandings of what is happening as they engage. Nor is this interaction most helpfully thought of as creating anything outside itself, such as a system with a boundary. Rather, global patterns of human relating

arise from many, many local interactions, paradoxically informing and informed by what both Bourdieu (1990) and Norbert Elias (2000) referred to as the *habitus*. The *habitus* is habitual and repetitive, but because it is dynamically and paradoxically emerging it also plays out in surprising, novel and sometimes unwanted ways because of the interweaving of intentions. Social life demonstrates the same characteristics of irregular irregularity which most intrigued the complexity scientists. Instead of developing abstractions to describe the emergence of both individuals and society, Stacey and colleagues are interested in figurations of power relations expressed in the dynamics of inclusion, exclusion and identity formation, communication understood as conversation, and evaluative choices which reveal ideology.

The implications of a radical view for evaluative practice

What would be the implications for evaluative practice for maintaining this more radical interpretation of the complexity sciences, and assuming that non-equilibrium and non-linearity were the default rather than the exception? Firstly, I think it would make highly problematic the idea that evaluators can decide whether social interventions are complex or not, or that they can be partially complex, no matter how simple they appear to be, and no matter how routine people's activities. Additionally, assuming a complex world would call into question the idea that emergence is a special category of social activity. Another way of understanding emergence is that social life is always emerging in one pattern or another, whether an intervention is tightly or loosely planned.

Next, there would be no reason to exclude the process of evaluation itself from an understanding that social life is predictably unpredictable. Evaluation is a situated, contextual practice undertaken by particular people with specific life-histories interacting with specific others, who are equally socially formed. The evaluative relationship is an expression of power relations, both between the commissioner of the social intervention/evaluation and the evaluator, and between both of these and the people comprising the intervention, which will inform how the evaluation emerges. One way of understanding the growing insistence of some commissioners of social interventions on highly abstract ways of representing and evaluating them is that it enables them to 'see like a state' in James C Scott's terms (1998), which I have explored elsewhere (Mowles, 2013). Scott argues that what he calls abstract and 'thin simplifications' cover over the very improvisational and adaptive activity that makes social projects work, and even improve them, and which

should be of interest both to commissioners and evaluators. Only a purist experimentalist assumes that the methods they are using have no impact on the people they are studying, or at least they believe they can control for them, but an evaluator convinced about complexity might also take an interest in how their own practice forms, and is formed by the relationships they are caught up in with the people they are evaluating. To take up Callaghan's invitation (2008), the evaluator would not just be interested in how people in the intervention negotiate order, but how the evaluation itself is negotiated. Equally, they would be interested in how power relations play out in, and affect, the social intervention.

Evaluators would cease hunting for mechanisms, would be less interested in logic diagrams, no matter how 'flexible', and would pay close attention to the quality of conversational life of social interventions, including how participants took up and understood any quantitative indicators that they might be using in the unfolding of the project. Evaluators might use narratives as a way of understanding how general intentions for social improvement play out in particular circumstances. Tsoukas (2009) argues that narrative is a good way of maintaining the paradoxical relationship between the particular and the general and a good example in the evaluative literature can be found in Greenhalgh et al.'s (2011) evaluative account of a large scale government IT project. They would also be interested in how the programme changed over time and how people accounted for these changes: 'progress' in terms of the social intervention, could also be understood in the movement of people's thinking and their sense of identity. Evaluators convinced of the importance of insights from the complexity sciences would argue, along with Sanderson (2009) that evaluators should assume a greater humility in their work and their claims about causality and replicability.

Concluding remarks

A wide variety of scholars have called orthodox evaluative practice into question and have argued for greater participation from beneficiaries, for greater thoughtfulness, responsiveness and humility on the part of commissioners and practitioners of evaluation, and for greater recognition of evaluation as a political activity, whether they mention the complexity sciences or not. In this article I have reviewed only those evaluation scholars who make a direct appeal to the complexity sciences in their work to offer a critique of what I consider to be some trends in the way they understand and take up the ideas.

I have argued that for the most part evaluation scholars try to accommodate insights from the complexity sciences in order to preserve the prevailing assumptions about the discipline of evaluation as a rational, stable activity somehow detached from the activities of the social intervention they seek to evaluate. I have argued, after Fleck (1979) that what I take to be some of the more radical insights then get subsumed within current orthodoxy. So, if the dominant view is that an evaluator takes a contingency approach to designing their evaluation, and that this is the prime task of the evaluator (Stern et al., 2012) then they must first decide whether the programme they are evaluating is complex or not, in whole or in part. If the social intervention is complex then particular techniques apply; if not, then not. Sometimes the best that a complexity perspective is thought to achieve is to make a propositional logic model more 'flexible'. Secondly, I have argued that sometimes scholars homogenise insights from the complexity sciences either to over-claim or to under-claim their importance. I have tried to point out how the complexity sciences have developed, and why that development makes a difference depending on which particular manifestation one chooses to engage with and interpret. It is important to deflate some claims, but challenge others: for example, the notion that emergent activity is somehow a special category of human relating and is the opposite of being tightly planned I would regard as a deflationary tendency in the evaluation scholarship. Thirdly I have tried to describe how what I take to be some of the more radical insights from the complexity sciences get lost in abstraction. For example, despite the claims of realistic evaluators to be interested in how and for whom social interventions 'work' and in generative causality, they still seem to be committed to finding underlying 'mechanisms' and regularities, always located at a different 'level' of reality.

As an alternative I have argued that if we were to assume that non-linearity in the social is the norm rather than the exception then this completely reframes the social activity of evaluation. In opening up the 'black box' of what people are doing to make social interventions work, we might become much more interested in fluctuating power relationships, including between the evaluator and the evaluated, in narrative accounts, and the paradoxical relationship between the general and the particular and how it is played out in many, many local interactions. We might accept that generalising from the 'success' or otherwise of particular social interventions will always be a probabilistic undertaking, and that there is no easy route to 'scaling up' or making unproblematic global assertions based on the perspective of 'seeing like a state'.

No funding has been received to support the writing of this article.

References

- Allen, P. (1998a) Evolving Complexity in Social Science, in Altman, G. and Koch, W.A. (eds) *Systems: New Paradigms for the Human Sciences*, New York: Walter de Gruyter.
- Allen, P. (1998b) Modelling Complex Economic Evolution, in Schweitzer, F. and Silverberg, G. (eds) *Selbstorganisation*, Berlin: Duncker and Humblot.
- Allen, P., Strathern, M. and Baldwin, J. (2006) Evolutionary Drive, New Understandings of Change in Evolutionary Systems, *Emergence: Complexity and Organisation*, vol 8, no 2.
- Aspen Institute (1997) *Voices from the Field: Learning from the Early Work of Comprehensive Community Initiatives*, Aspen Institute, Washington, DC.
- Banerjee, A. and Duflo, E. (2012) *Poor Economics: Barefoot Hedge-fund Managers, DIY Doctors and the Surprising Truth about Life on less than \$1 a Day*, London: Penguin.
- Barnes, M., Matka, E. and Sullivan, H. (2003) Evidence, Understanding and Complexity in Non-linear Systems, *Evaluation*, 265-284.
- Bhaskar, R. (1978) *A Realist Theory of Science*, 2nd Edition, Brighton: Harvester Press.
- Bourdieu, P. (1990) *The Logic of Practice*, Cambridge: Polity.
- Callaghan, G. (2008) Evaluation and Negotiated Order: Developing the Application of Complexity Theory, *Evaluation*, Vol 14 (4), 399-411.
- Chen, H. and Rossi, P. (1989) Issues in the Theory-driven Perspective, *Evaluation and Programme Planning*, 12(4): 299-306.
- Chia, R. (1998) From Complexity Science to Complex Thinking: Organization as Simple Location. *Organization*, 5(3): 341–370.
- Connell, A., Kubisch, C., Schorr, L.B. and Weiss (eds) (1995) *New Approaches to Evaluating Community Initiatives: Vol 1, Concepts, Methods and Contexts*, Washington DC, The Aspen Institute.
- Davies, R. (2004) Scale, Complexity and the Representation of Theories of Change, *Evaluation*, Vol 10(1): 101–121.
- Davies, R. (2005) Scale, Complexity and the Representation of Theories of Change Part II, *Evaluation*, Vol 11(2): 133–149.
- De Marchi, S. and Page, S.E. (2008) Agent-based Modelling, in Box-Steffenmeier, J.M., Brady, H. E. and Collier, D. (eds) *The Oxford Handbook of Political Methodology*, Oxford: Oxford University Press.
- Dewey, J. (2005) *The Quest for Certainty: A Study of the Relation of Knowledge and Action*, New York: Kessinger Publishing.
- Elias, N. (1939/2000) *The Civilising Process*, Oxford: Blackwell.

- Fleck, L. (1979) *Genesis and Development of a Scientific Fact*, Chicago: University of Chicago Press.
- Forss, K. and Schwartz, R. (2011) Introduction, in Forss, K., Marra, M. and Schwartz, R. (eds) *Evaluating the Complex: Attribution, Contribution and Beyond: Comparative Policy Evaluation vol 18*, New Brunswick: Transaction Publishers.
- Forss, K., Marra, M. and Schwartz, R. (eds) (2011) *Evaluating the Complex: Attribution, Contribution and Beyond: Comparative Policy Evaluation vol 18*, New Brunswick: Transaction Publishers.
- Funnell, S. and Rogers, P. (2011) *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*, San Fransisco: Jossey-Bass.
- Glouberman, S. and Zimmerman, B. (2002) Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like?, Discussion Paper No. 8: Commission on the Future of Health Care in Canada, Plexus Institute.
http://www.plexusinstitute.org/resource/collection/6528ED29-9907-4BC7-8D00-8DC907679FED/ComplicatedAndComplexSystems-ZimmermanReport_Medicare_reform.pdf.
- Greenhalgh, T., Russell, J., Ashcroft, R. and Parsons, W. (2011) Why national ehealth programs need dead philosophers: Wittgensteinian reflections on policymakers' reluctance to learn from history, *The Millbank Review*, 89.
- Hedström, P. (2005) *Dissecting the Social: on the Principles of Analytical Sociology*, Cambridge: Cambridge University Press.
- Hedström, P. (2008) Studying Mechanisms to Strengthen Causal Inferences in Quantitative Research, in Box-Steffenmeier, J.M., Brady, H. E. and Collier, D. (eds) *The Oxford Handbook of Political Methodology*, Oxford: Oxford University Press.
- Kuhn, T. (1996) *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press.
- Marra, M. (2011) Some Insights from Complexity Science for the Evaluation of Complex Policies, in Forss, K., Marra, M. and Schwartz, R. (2011) *Evaluating the Complex: Attribution, Contribution and Beyond: Comparative Policy Evaluation vol 18*, New Brunswick: Transaction Publishers.
- Mead, G.H. (1934) *Mind, Self and Society from the Standpoint of a Social Behaviourist*, Chicago: University of Chicago Press.
- Miller, J. and Page, S.E. (2007) *Complex Adaptive Systems: an Introduction to Computational Models of Social Life*, Princeton, Princeton University Press.
- Mowles, C., Stacey R. & Griffin, D. (2008) What contribution can insights from the complexity sciences make to the theory and practice of development management? *Journal of International Development*, Volume 20, Issue 6, 804-820.
- Mowles, C. (2011) *Rethinking Management: Radical Insights from the Complexity Sciences*, London: Gower.

Mowles, C. (2013 forthcoming) Evaluation, complexity, uncertainty: theories of change and some alternatives in Wallace, T., Porter, F. and Ralph-Bowman, M. (eds) *Aid, NGOs and the reality of women's lives: a perfect storm*, Rugby: Practical Action Publications.

Nagel, T. (1986) *The View from Nowhere*, Oxford: Oxford University Press.

Patton, M.Q. (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York: The Guilford Press.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, London: Sage.

Pawson, R. (2006) *Evidence-based Policy*, London: Sage.

Prigogine, I., and Nicolis, G. (1977) *Self-Organization in Non-Equilibrium Systems*, London: Wiley.

Rogers, P. (2008) Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions, *Evaluation*, Vol 14 (1), 29-48.

Ray, T.S. (1992) An Approach to the Synthesis of Life, in Langton, G., Taylor, C., Doyne Farmer, J. and Rasmussen, S. (eds) *Artificial Life II*, Santa Fe Institute, Studies in the Sources of Complexity, vol 10, Reading: Addison Wesley.

Reynolds, C. W. (1987) Flocks, Herds and Schools: a Distributed Behaviour Model, *Proceedings of Siggraph '87, Computer Graphics*, vol 21, no4, 25-34.

Rogers, P. (2011) Implications of Complicated and Complex Characteristics for Key Tasks in Evaluation, in Forss, K., Marra, M. and Schwartz, R. (2011) *Evaluating the Complex: Attribution, Contribution and Beyond: Comparative Policy Evaluation vol 18*, New Brunswick: Transaction Publishers.

Sanderson, I. (2000) Evaluation in Complex Policy Systems, *Evaluation*, 6 (4), 433–454.

Sanderson, I. (2009) Intelligent Policy Making for a Complex World: Pragmatism, Evidence and Learning, *Political Studies*: vol 57, 699–719.

Scott, J. C. (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*, Yale University Press, New Haven, CT.

Senge, P, Scharmer, C.O., Jaworski, J. and Flowers, B.S. (2005) *Presence: Exploring Profound Change in People, Organisations and Society*, London: Nicholas Brealey Publishing.

Smith, M.F. (1989) *Evaluability Assessment: A Practical Approach*, Boston: Kluwer Academic Publishers.

Spicer, N. and Smith, P. (2008) Evaluating Complex, Area-Based Initiatives in a Context of Change: the Experience of the Children's Fund Initiative, *Evaluation*, Vol 14 (1), 75-90.

Spinatsch, M. (2011) Evaluating a Complex Policy in a Complex Context: the Elusive Success of the Swiss Smoking Prevention Policy, in Forss, K., Marra, M. and Schwartz, R. (eds) *Evaluating the Complex: Attribution, Contribution and Beyond: Comparative Policy Evaluation vol 18*, New Brunswick: Transaction Publishers.

- Stacey R (1992) *Managing the Unknowable*, San Francisco: Jossey-Bass.
- Stacey, R. (2011) *Strategic Management and Organizational Dynamics: the Challenge of Complexity*, 6th Edition, London: Prentice Hall.
- Stacey, R. (2012) *Tools and Techniques of Leadership and Management: Meeting the Challenge of Complexity*, London: Routledge.
- Stame, N. (2004) Theory-based evaluation and types of complexity, *Evaluation*, 10: 58, 58-76.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*, Working Paper 38, Department for International Development, London.
- Stern, E. (2008) Editorial, *Evaluation*, Vol 14 (1), 5-7.
- Taylor, C. (1999) To Follow a Rule in Shusterman, R. (ed.) *Bourdieu: a Critical Reader*, Oxford: Blackwell.
- Tsoukas, H., and Hatch, MJ. (2001) Complex thinking, complex practice: the case for a narrative approach to organizational complexity, *Human Relations*, 54(8): 879–1013.
- Tsoukas, H. (2006) *Complex Knowledge: Studies in Organizational Epistemology*, Oxford University Press: Oxford.
- Tsoukas, H. (2009) Craving for Generality and Small-n Studies: A Wittgensteinian Approach towards the Epistemology of the Particular in Organization and Management Studies, in *The Sage Handbook of Organizational Research Methods*, ed. D.A. Buchanan and A. Bryman, 285–301. London: Sage.
- Van der Knaap, P. (2011) Sense and Complexity: Initiatives in Response to Performance Audits, *Evaluation*, 17: 4, 351-363.
- Weiss, C. (1995) Nothing so practical as a good theory: Exploring Theory-based evaluation for Comprehensive Community Initiatives for Children and Families, in Connell, A., Kubisch, C., Schorr, L.B. and Weiss (eds) *New Approaches to Evaluating Community Initiatives: Vol 1, Concepts, Methods and Contexts*, Washington DC, The Aspen Institute.
- Weiss, C. (1997) Theory-based Evaluation: Past, Present and Future, in Rog, D.J. *Progress and Future Directions in Evaluation, New Directions for Evaluation*, 76, San Fransisco: Jossey Bass.
- Westhorp, G. (2012) Using Complexity-consistent Theory for Evaluating Complex Systems, *Evaluation*, 18 (4), 405-420.