

(Breaking) The Iron Triangle of Evaluation*

Martin Reynolds

Abstract Ideas from complexity science and systems thinking are demonstrably helpful in a shift from exploring (systematic) linear net effects of an intervention towards exploring wider (systemic) effects occurring elsewhere. But where these ideas of ‘impact’ are coupled with a narrow use of the contingency approach, some less helpful ‘triangulated’ relationships might be evident. These relationships might be regarded in terms of an ‘iron triangle’, a metaphor used frequently to exemplify pernicious relations of power. The most notable expression of the iron triangle is the ‘military–industrial complex’. This article briefly outlines generic features of the iron triangle in terms of ‘systemic triangulation’ – an idea linking three core systems concepts of interrelationships, perspectives and boundaries. Drawing on a tradition of systems thinking in practice, an associated systemic triangulator is introduced as both a diagnostic and planning heuristic; a device for not only diagnosing symptoms of an evaluation–industrial complex but for prompting ideas towards a more benign evaluation–adaptive complex for impact evaluation.

1 Introduction

I use it [systems thinking] often. But I can see why people surrounding me don't use it often. They think it's abstract. It's not solving problems the way they would like...

Gets a bit atomistic with people proselytising their favourite words and models and flogging old horses with new names, losing its pluralistic origins...

The statements come from some free text responses on a small (though global) online survey (86 respondents) amongst professionals with knowledge of systems thinking in different domains (Lachica 2013). Despite a clear need for countering prevailing reductionist approaches to impact evaluation, the comments flag issues regarding the take-up of systems ideas in evaluation practice. One issue is a reluctance to emphasise the systems idea as indeed an abstract idea – as though abstracting from reality is somehow neither normal nor helpful. A second issue is that systems thinking is often regarded as something quite separate from, rather than potentially integral to, other practices. Both issues can lend themselves towards maintaining a model of business-as-usual in evaluation thinking in practice.

Formulating evaluative questions of impact from real-world complexities requires a process of abstraction. The problem of impact evaluation as described through articles in this *IDS Bulletin* might be summarised in terms of a broadening of the questions to be addressed; from concerns of impact in narrow terms of outputs to concerns about wider outcomes being effected. A shift in attention amongst evaluators is generally acknowledged; from exploring questions regarding the linear net effect of an intervention (project, programme or policy) towards questions regarding whether an intervention made a difference to the situation, what the difference was, how the difference was made, and what differences might have been made elsewhere. It represents a shift from being systematic towards being more systemic. Such a shift in questioning actually involves going up a level of abstraction in order to deal with the complexities of interventions.

Flood and Carson (1993) suggested that the purpose of applied systems thinking is to deal with situations of complexity. A cherished hallmark of good systems thinking is the aspiration to make simple the complex. Such an aspiration may not sit comfortably amongst

those who sometimes confuse systems with the reality of a situation; or as the age-old systems adage proclaims – those who confuse the map for the territory (Korzybski 1995 [1933]). The downside of an aspiration to simplify is a rebuke that systems thinking can be too abstract. At one level the rebuke is well founded. When a system becomes too divorced from the realities that it needs to address, its reliability can be questioned. There is clearly something wrong when, say, a health system assumes that sickness can be fixed solely by increasing medication, or a penal system is assumed to work purely on a basis of punishing offenders, or an economic system assumes profit motive to be the sole driver of economic success. Likewise, when systems of impact evaluation are simplified to the extent of assuming a reality of linear causality with independent parts – such as with some extreme versions of randomised controlled trials (RCTs) – then the impacts measured are likely to be over-simplified and hence bear little sense of resemblance to reality.

But let us be clear and unambiguous, systems thinking is an explicit endeavour of abstracting from reality. The challenge is to generate abstracts that have not just reliability towards mapping realities of the real world, but resonance and relevance to the situations being addressed. So rather than having systems of impact evaluation based on, say, simple abstractions of linear, mechanistic cause-and-effect, it may be more appropriate to go up a further level of abstraction – mirroring the shift from systematic to systemic impact questioning noted previously – for example, to consider systems with parts interrelated in non-linear ways, creating feedback loops, tipping points, path dependencies, time delays, and so on, all contributing to impacts that may not be so easily predicted. The contributions to this issue of the *IDS Bulletin* provide different means of using this understanding of reality. Many draw on the helpful idea of reality abstracted as a complex adaptive system (CAS) (*cf.* Patton 2010; Forss, Marra and Schwartz 2011; Ramalingam 2013).

A key feature shared by many authors in the use of CAS is the assumption that not all situations can be identified as complex, and therefore not all situations demonstrate features of a complex adaptive system. There has emerged a ritualised practice in evaluation with supposing that

situations can be mapped from the outset into being either simple, complicated, or complex – an idea initiated by Ralph Stacey in terms of organisational management (Stacey 1992) but later adopted in the evaluation field (*cf.* Glouberman and Zimmerman 2002). Such thinking invokes a contingency approach whereby the type of method adopted for an evaluation is contingent upon the type of situation. A popular metaphoric expression of such situations is in distinguishing between three different activities: (i) baking a cake; (ii) sending a rocket to the moon; and (iii) bringing up children. Whereas the former two activities might respectively be regarded as being ‘simple’ (following recipe instructions) and ‘complicated’ (requiring address to many variables, but nevertheless predictable in outcomes), it is only the latter skill in parenting that invokes ‘complexity’ where the situation is volatile, uncertain and highly unpredictable.

Chris Mowles (2014) provides an insightful critique of contingency theory in relation to the use of complexity ideas in evaluation. His core argument draws explicitly on Stacey’s later rejection of his own tripartite heuristic (Stacey 2001) by further suggesting that it might be better to assume that *all* situations are complex. Situations comprise humans and human interaction is always complex and emergent. A systems thinking perspective might go further and suggest that all situations consist of interrelationships that may be regarded at any one time as being both (i) complicated – interconnected – as evidenced in, for example, chaos theory, and (ii) complex – involving people, even if only as mere observers, and therefore inviting multiple perspectives. Furthermore, a *critical* systems thinking perspective might regard all situations as (iii) conflictual – with differences in relations of power circumscribing what is deemed to be important with regard to boundaries of interconnections (what’s in and what’s out) and boundaries of perspectives (whose viewpoints are more privileged and whose viewpoints are more marginalised) (Reynolds 2011).

There are two points of departure here from contingency theory. Firstly, as with Mowles, with systems thinking in practice we would argue that there are no actual ‘simple’ situations, given that ‘simplicity’ represents a particular *perspective on a situation* rather than ‘the’ actual *attribute of the*

situation. An artefact from a situation such as a documented project, programme or policy may of itself be regarded as ‘an’ attribute. As ‘an’ artefact it may be regarded as simple or even complicated (depending on the relative number of variables – including different perspectives – being dealt with). But the artefact in and of itself is not complex. The complexity arises in the auditing, planning, implementing and evaluating of the intervention – that is, where and when different perspectives are actually at play. Secondly, whilst any situation might helpfully be regarded as comprising all three interlocking features of complicatedness (through interrelationships), complexity (through multiple perspectives) and conflict (through boundary judgements), there remains a question on whether it is helpful to delineate these situational features in order to align particular evaluative methods – the ‘horses-for-courses’ thinking underpinning much of contingency theory. Whilst there are now a vast array of different evaluative methods, including different systems and complexity tools, it is perhaps helpful to remind ourselves that such tools are conceptual devices rather than reified objects; and hence open to change, adaptation and innovation depending on the user and context of use. Effectiveness may not necessarily be contingent on ‘the method’ and the ability to adopt new tools and associated language, but rather the user’s experience and craft skills in dealing with all three features of an evaluand – complicatedness, complexity and conflict – using whatever tools with which s/he may have experience.

The distinction between situations and systems is key to diagnosing potential problems of contingency thinking and the helpfulness of complexity thinking in impact evaluation, whilst avoiding the pitfalls of seemingly being too abstract and/or too atomistic. Situations might be referred to as ‘messes’ as distinct from ‘systems of interest’ (Ackoff 1974), or as ‘wicked problems’ – or even ‘super-wicked problems’ such as with climate change (Levin *et al.* 2012) – as distinct from ‘tame problems’ (Rittel and Webber 1973). Systems thinking is therefore a process of making simple the complex; a process of abstracting from reality.

This article will take up a further level of abstraction. Whereas a complex adaptive system can provide a helpful abstract proxy to individual

interventions being evaluated, a similar systems idea can be used as an abstract heuristic for understanding the wider set of influences associated with the political economy of evaluation (see also Winckler Andersen in *IDS Bulletin* 45.6, 2014). The simplicity of the system captures rather than detracts from the reality of the evaluation process. The proposed device is based on ideas of ‘systemic triangulation’; a generic device that we might call a systemic triangulator.

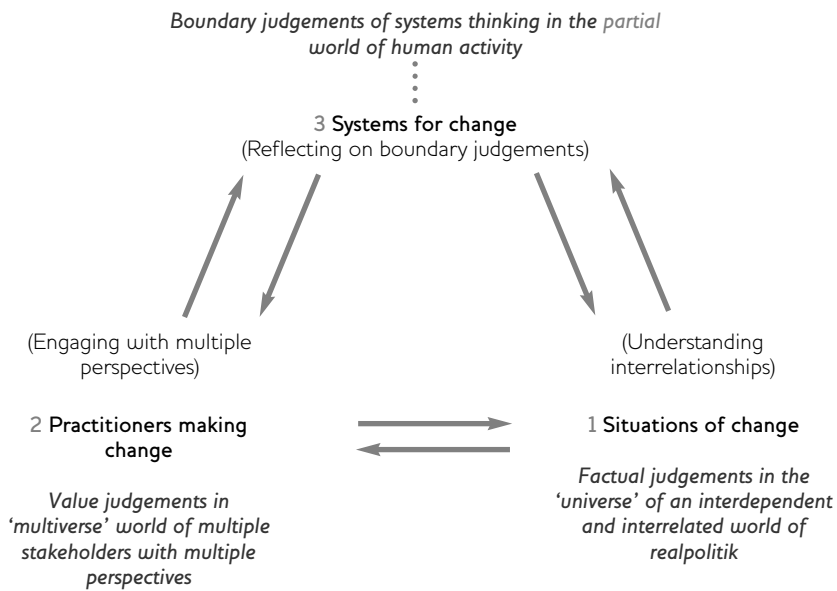
After briefly introducing the idea of systemic triangulation and considering its relationship with the iron triangle metaphor, the political economy of evaluation expressed in terms of an evaluation–industrial complex will be explored. By using systems thinking in practice some questions are prompted regarding the influences between an evaluand, evaluators and evaluations with particular attention to contemporary impact evaluation. The article finishes with some thoughts on what a more benign system of impact evaluation might look like.

2 From systems concepts to systemic triangulation

The three core concepts associated with systems thinking – interrelationships, (multiple) perspectives and boundary (judgements) – mark a significant improvement in understanding and engaging with essential features of systems thinking. These are generally regarded as standalone entities which may resonate with entities of value and learning (Hummelbrunner and Reynolds 2013; see also article by Hummelbrunner in this *IDS Bulletin*). But these entities might also be transformed into an active endeavour – a heuristic of systems thinking in practice – comprising verbs as well as nouns; *understanding* interrelationships, *engaging* with multiple perspectives and *reflecting* on boundary judgements (Reynolds 2011, 2013; Williams, this *IDS Bulletin*). In this article the heuristic is adapted further to provide a template for evaluating the political economy of evaluation processes, something that was also suggested by Winckler Andersen in *IDS Bulletin* 45.6 (Winckler Andersen 2014). To do this I will draw on one of the most profound generic expressions of systems thinking – the notion of systemic triangulation developed by Werner Ulrich (1998, 2000, 2003).

Systemic triangulation represents an extension of ‘triangulation’ familiar in the empirical social sciences. The extension is to rely not only on

Figure 1 The systemic triangulator



Source Adapted from Reynolds (2013) and the original 'eternal triangle' of Ulrich (2000: 252).

different methods and theories for understanding reality (interrelationships), but also for engaging with alternative value sets (perspectives) and reflecting on reference systems (boundaries) as a source of assessing real-world practice (cf. Ulrich 2003: 334). Systemic triangulation refers to what Ulrich calls an 'eternal triangle' of interdependence between judgements of 'fact', value judgements and boundary judgements (Ulrich 1998, 2000):

Thinking through the triangle means to consider each of its corners in the light of the other two. For example, what new facts become relevant if we expand the boundaries of the reference system or modify our value judgments? How do our valuations look if we consider new facts that refer to a modified reference system? In what way may our reference system fail to do justice to the perspective of different stakeholder groups? Any claim that does not reflect on the underpinning 'triangle' of boundary judgments, judgments of facts, and value judgments, risks claiming too much, by not disclosing its built-in selectivity (Ulrich 2003: 334).

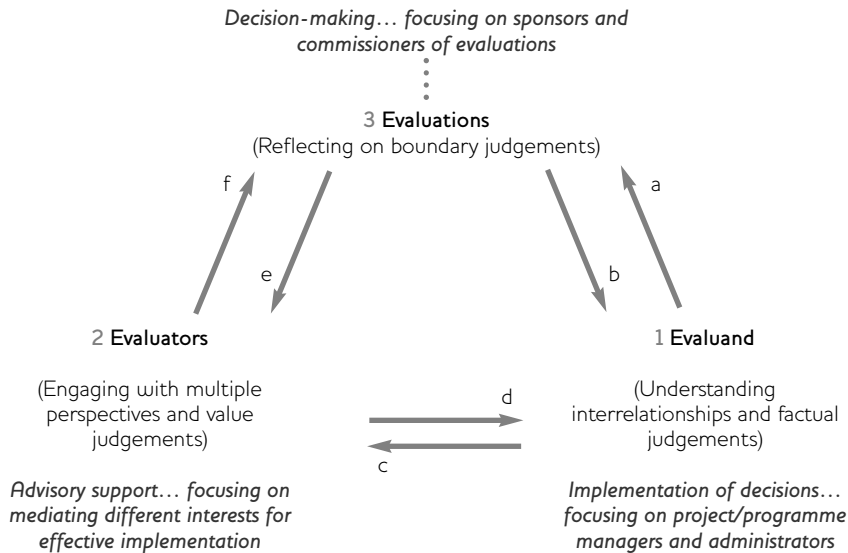
Figure 1 illustrates the influences between these three entities. The systemic triangulator builds on Ulrich's original figurative triangle of facts,

values and boundary judgements (Ulrich 2000: 252). In my rendition of the triangle, 'facts' relate to the situations of change in the real world associated with infinite *interrelationships*, 'values' relate to practitioners managing change, associated with the core idea of multiple *perspectives*, and 'boundaries' relate to the ideas/tools/models used for decision-making about managing change, associated with the core idea of reflecting on *boundary* judgements.

Making boundary judgements is, as Ulrich points out, a 'partial' activity in two senses of the word: (i) partial in representing only a section rather than the whole of the total universe of interrelationships (what's in and what's out?); and (ii) partial in serving some stakeholder perspectives or interests – including practitioners' interests – better than others (whose views are privileged and whose are marginalised). The inevitable partiality in systems thinking makes clear the critical limits of being 'holistic' – getting the big picture – (cf. Williams, this *IDS Bulletin*) and of being 'pluralistic' – being inclusive of stakeholders whilst claiming an unbiased view or value judgement on different perspectives (cf. Hummelbrunner, this *IDS Bulletin*).

The idea of systemic triangulation is itself strongly influenced by traditions of critical social

Figure 2 Evaluation-in-practice represented as possibly an evaluation–industrial complex: an influence diagram illustrating six activities – (a) auditing; (b) planning; (c) evaluating (understanding); (d) evaluating (practising); (e) commissioning; and (f) learning



Source Author's own.

theory – particularly Jurgen Habermas – and American pragmatism – particularly the works of Charles Peirce, William James, John Dewey, Edward Singer and C. West Churchman (see Ulrich 1983 for a detailed scholarly account of an earlier synthesis of these ideas). The remainder

of this article explores the resonance that systemic triangulation may have on the metaphor of the iron triangle used in political science, and how this metaphor might be helpful to explore systemic failures generally, and systemic failure in (impact) evaluation practice more specifically.

Table 1 Expressions of the iron triangle compared

The iron triangle: three generic dimensions	(original) Paris Peace Treaty (1919)	Military–industrial complex (1961)	Evaluation–industrial complex (2014)
i Interrelationships (contexts)	(on the ground) professional soldiers (militarism)	Bureaucracy	Evaluand... including managers/administrators implementing interventions*
ii Multiple perspectives (people)	Interests of military industry (materialism)	Interest groups	Evaluators... with the task of representing (and inevitably) privileging different interests
iii Boundary judgements (ideas/tools)	Decisions of politicians (bourbonism)**	Decisions of Congress	Evaluations... as expressions of advice to decision-makers (commissioners of interventions)

*Interventions can include any intended action on the real world, ranging from simple plans or projects, to wider programmes or policies, and from local to global levels.

**Bourbonism depicts extreme political and social conservatism. It draws on subservience to support for the rule of the Bourbons, a European royal line that ruled in France from 1589–1793 and 1815–48, and in Spain (1700–1808; 1813–1931) and Naples and Sicily (1734–1806; 1815–1860).¹

Source Author's own.

Table 2 Two contrasting views of evaluation-in-practice

Activities	Evaluation–industrial complex	Evaluation–adaptive complex
a Auditing – overview the site of implementing an intervention (project, programme, policy, plan...)	Situations of an evaluand are <i>systematically</i> recognised at outset as either tame (simple or complicated with linear cause-and-effect attributes), or wicked (complex with non-linear, unpredictable effects and impacts).	Situations of an evaluand are <i>systemically</i> viewed as comprising all of (i) complicatedness (interrelationships), (ii) complexity (multiple perspectives), and (iii) conflict (contrasting boundary judgements). Side-effects are inevitable.
b Planning – design of terms of reference (ToR) for the intervention	Planning is <i>purposive</i> – based on a perceived need for rather fixed goals and targets. Where need be, with ‘complex’ situations, a dominant imperative is to ‘tame’ any ‘wicked-ness’ in the situations depending, for example, on quantitative measures of success. Little room for flexibility and adjusting measures over time.	Planning is <i>purposeful</i> – based on ‘agile’ measures of success that are not rigidly defined by fixed quantitative indices alone, but also qualitative indices (stories, narratives, case studies, testimonies, etc.); both adaptable to changing situations in the evaluand and hence being flexible, adaptive and imaginative.
c Evaluating 1: understanding	Evaluator unquestionably adopts measures of success passed on through ToR of the evaluand. Regards evaluand as a fixed operational system (ontological mechanistic reality). Adopts disinterested (quasi) <i>positivist epistemology</i> involving as much as possible objective valuation with prime attention to avoiding bias. Focus more on efficiency in relation to efficacy and effectiveness.	Evaluator able to question ToR of evaluand, including its own measures of success. Regards evaluand as a learning system adapting to changing circumstances. Adopts radical <i>constructivist epistemology</i> involving self-awareness of biases in shaping the evaluation. Able to continually juggle between focus on efficacy, efficiency and effectiveness (including ethical and political notions of equity and sustainability).
d Evaluating 2: practice	Practitioner role as <i>measuring</i> of value, either as specialist user of same tools – ‘bestpracticitis’, or seeking an ever growing ‘toolbox’ <i>adopting</i> new tools as ‘best fit’ for purpose (‘horses-for-courses’).	Practitioner regarded as an agile ‘ <i>bricoleur</i> ’, a craftsman formatively <i>developing</i> value (instrumental, intrinsic and critical) through <i>adapting</i> tools with new ideas during the course of an evaluation. /cont.

3 The iron triangle and an evaluation–industrial complex

The iron triangle metaphor was first expressed by Ralph Pulitzer, a political journalist reporting critically on the Paris Peace Conference amongst victorious allied governments following the First World War in 1919. Pulitzer warned against the insipid confluence of interests amongst three exclusive sets of actors – the military, industry and politicians making the decisions (Pulitzer and Grasty 1919). The iron triangle is helpful in speaking to a wider sense of systemic failure; one popular expression being the ‘military–industrial complex’ used by American President Dwight Eisenhower during his 1961 presidential resignation speech. Political activists like Angela

Davis and Arundhati Roy have adapted ideas of ‘industrial complex’ and ‘iron triangle’ respectively to surface pernicious confluences of interest in more specific domains. Similarly, radical economists like J.K. Galbraith amongst other academics have used and developed these ideas in studies of political economy. Elsewhere, I have used the iron triangle metaphor to analyse systemic failure amongst academic economists in the midst of the global economic crisis of 2008 (Reynolds 2014).

The metaphor is generically used to describe interaction between three entities: (i) some loosely defined ‘bureaucratic’ entity which represents the site of real-world implementation

Table 2 Two contrasting views of evaluation-in-practice (cont.)

Activities	Evaluation–industrial complex	Evaluation–adaptive complex
e Commissioning ToR for evaluators of interventions requiring some guarantee or assurance of trustworthiness	Rigour narrowly defined in terms of objective <i>reliable and replicable</i> use of evaluation tools as fixed commodities; proven through application in case studies understood as being suitable for <i>either</i> simple/complicated (tame) or complex (wicked). Assurances relating to resonance with existing practices tend not to be given importance. Rigour associated with demonstrating responsibility is limited to providing assurances of accountability to decision-makers rather than serving wider social and ecological imperatives.	Rigour understood as three sets of co-guarantor attributes: (i) <i>transparency of limitations</i> in different contexts, rather than some claim towards achieving ‘objective truth’; (ii) <i>complementarity or resonance</i> – communicability with other disciplines/groups/cultures, etc.; and (iii) <i>social and ecological responsibility or relevance</i> – internal coherence in dealing with <i>all</i> inevitable issues of complicatedness, complexity and conflict, and external coherence in addressing issues of social and ecological responsibility.
f Learning... gained by decision-makers from evaluations	Enables single-loop and sometimes double-loop learning, but generally less reflective of power relations circumscribing the primary intervention and/or circumscribing the secondary intervention in terms of evaluating the intervention. Evaluation regarded as apolitical.	Enables single-loop, double-loop and triple-loop learning from an evaluation of an evaluand, including learning that all evaluands have degrees of complexity and conflict. Evaluation regarded as political, but also involving humility in dealing with uncertainty.

Source Author’s own.

of decisions (e.g. civil servants, managers and administrators); (ii) interest groups/individuals who stand to benefit from the implementation of decisions (e.g. commercial and corporate interests of various kinds, or commissioned advisory groups whose task is to capture different interests); and (iii) decision-makers themselves responsible for making and justifying decisions (e.g. Congress or Parliament, or at a lower level, commissioners of interventions).

For the purpose of exploring the industry of evaluation, the iron triangle metaphor might be used to describe interaction between three domains: (i) some site of real-world impact of an intervention – project, programme or policy – with managers and administrators responsible for implementing the intervention; (ii) evaluators as an interest group active in the process of evaluating interventions – principally and ideally capturing different values from different perspectives; and (iii) decision-makers – commissioners, foundations, donor agencies – ultimately responsible for interventions with implicit and/or explicit models of decision-making to justify actions taken, including the commissioning of evaluations. Table 1 maps out

the correspondence between different expressions of the iron triangle.

Figure 2 is a broad representation of the triadic relationships between these three domains for what might be called ‘evaluation-in-practice’ and which may have particular expression through an evaluation–industrial complex.

The influence diagram might be used as a diagnostic systemic tool for identifying flaws and opportunities associated with evaluation – whether impact evaluation or any other formalised set of evaluation practices, and at any level of intervention – whether evaluating a policy, programme or project directly, or evaluating an evaluation as an intervention in itself.

The notes below give some guidance on how the arrows of influence may contribute towards systemic failure of an evaluation–industrial complex. Although the notes attempt to focus on impact evaluation, they may be deemed relevant to evaluation more generically, particularly evaluation based on a contingency approach. Readers may find reference to Table 2 helpful for further understanding of activities (a) – (f).

a **Auditing situations** A proposed intervention involves some initial auditing in making (boundary) judgements associated with the factual reality of the intervention situation (the evaluand). A first audit based on contingency thinking would determine whether the situation (evaluand) is one where the interrelationships might be bounded as simple or complicated or more complex. Is the situation akin to invoking a quick fix solution – like recipe reading in baking a cake? Or is it one requiring lots of logistical support – like sending a rocket to the moon? Or is it very unpredictable and uncertain – like parenting a child? Many initial audits to guide interventions are premised on a false pre-supposition that the problem situation can be delineated at the outset between being tame or wicked... between being a mere difficulty or a mess... between being merely complicated or complex. The auditing ignores the idea that people – with different perspectives – make all social situations not only complex but often conflictual.

The key task for decision-makers responsible for initiating interventions² is to make initial boundary judgements that appear to be ‘true’ to the reality of the situation. The problem though is in assuming that such boundary judgements based on some ‘truth’ of the situation are able to be made from the outset. For example, a problem situation might be defined in terms of the prevalence of measles in children within a national population. Eradicating measles amongst the under-fives might be regarded as a relatively tame problem – simple or possibly, complicated, but not necessarily ‘complex’. However, important caveats might be made in regarding this as a ‘tame’ rather than a ‘wicked’ problem. It may only be regarded as tame *if* important ‘inputs’ or conditions are given – including for example, availability of effective medications, trained staff and management protocols, etc. Clearly, the risk is in auditing ‘situations’ as though they are ‘given’ systems – viewing messy situations as though they are just implicitly well-defined problems – confusing the territory for the map. There is an error here in factual judgement which regards the real world as ‘simple’ or ‘complicated’ (i.e. tame) without appreciating the human factor that makes *all* social situations complex and

potentially conflictual (i.e. wicked). It is human agents – people – that translate ‘wicked’ situations into ‘tame’ situations – humans that translate messy situations into systems. The eradication of measles amongst under-fives is as complex as preventing the spread of Ebola; it is a situation in which wider factual judgements of context regarding culture, history, geography and politics clearly have a significant role in shifting the problem from being ‘tame’ to being more ‘wicked’.

Might it be more useful then to assume complexity from the outset, given that complexity is determined not just by complicatedness of there being many interrelated variables, but by the human factor of there being multiple perspectives on the situation? Moreover, the incidence of multiple perspectives prompts alertness to there being actual and potential conflicts – conflicts in contrasting boundary judgements. A more useful principle might be to start any audit more systemically, taking account of not just interrelated variables in the situation, but different perspectives and potential conflicts amongst key stakeholders.

b **Planning (terms of reference) interventions** A typical sequence of tasks in planning an intervention involves firstly establishing an overall purpose, then constructing or designing reasonable objectives, and then defining some measures of success by which the intervention might then be evaluated. Defining the purpose represents the prime boundary of the system of intervention being initially designed. Two questions arise. First, at what level of planning are the purposes or objectives located – policy design, programme/project management, or at the level of operational administration? Second, how much flexibility is there in the system – do the purposes/objectives/measures remain fixed or are they open to change? For example, regarding the planning involved with confronting an issue of disease prevention, is it just a problem of ‘administering’ vaccinations at an operational level of planning, or might it be a set of objectives at a higher level of planning, possibly involving new management structures or even a change in national health policy? However initially perceived, it is helpful for evaluators of an intervention to regard these objectives in a wider contextual field as

nested purposes or objectives; that is, circumscribed by a supra-level system of purpose or worldview, and drivers of implied sub-level systems of more tactical ‘targets’. Regarding purposes in this way enables evaluators to see the intervention as being *purposeful* rather than *purposive* (see also Hummelbrunner in this *IDS Bulletin*). A purposive system is one where the purposes remain rigid and unfluctuating, more commonly associated with mechanically designed systems like washing machines and watches. Human systems are, or ought to be, purposeful in the sense that purposes are fluid and subject to change given inevitable change in circumstances (Ackoff and Emery 1972). A key problem shared amongst many evaluators is that they see little room for manoeuvre in advising against proffered objectives of an intervention (see also Williams in this *IDS Bulletin*).

‘Measures of success’ against which to monitor and evaluate ‘impact’ of intervention objectives are themselves boundary judgements. Ideally such measures ought therefore not to be fixed. A common critique of the most prevalent planning tool in international development for surfacing such measures – logical framework analysis (LFA) – is the assumed inflexibility and rigidity. The LFA approach tracks a pathway from resources inputs to activities, outputs, outcomes and impacts in a sequential unilateral manner helpful at the operational/administrative level of planning, where clear measures in terms of performance indicators can be identified. But it has less value for complex situations. When applied to such situations, the messy realities can become dangerously entrenched as being simple and subject to linear cause-and-effect explanations.

- c **Evaluating 1: Understanding (interrelationships)** Two aspects of evaluation might usefully be conceptually demarcated – firstly, *understanding* the context of the evaluand, and secondly, engaging in *practice* with the various agencies of the evaluand. In real-world praxis the two are, of course, intricately interwoven; as with corresponding dualities of theory/practice, thinking/action, as well as the thinking in practice of summative/formative evaluation. Abstracting out understanding and practice can help reveal the more direct interwoven

sources of influence on evaluators’ praxis. Item (d) will explore further the influences on evaluation practice – a more formative endeavour of developing value – whilst item (c) here will focus on how understandings of value in the context are influenced (see Williams, this *IDS Bulletin*).

Evaluators are essentially tasked with translating factual judgements relating to the real-world complicatedness, complexity and conflict, into appropriate value judgements. Measures of success – suggested by the intervention’s terms of reference (ToR) – might be translated in terms of efficacy and efficiency (has the intervention worked according to its mandate with good use of resources?) and effectiveness (what wider effects may have been generated by the intervention?). As Hummelbrunner points out (this *IDS Bulletin*), a third concern for evaluators might also be to understand wider critical and political values of equity and sustainability in relation to the intervention. The task then is to capture these realities of an intervention in terms of appropriate values – instrumental, intrinsic and critical/political. Practitioners in the evaluation industry may like to consider the extent to which their evaluations are biased towards the instrumental values passed down through the ToR.

Evaluation is often regarded normatively as being a bias-free pursuit of ‘truth’. The task traditionally is to capture the realities of interrelationships in as an objective (and hence ironically, value-free) way as possible using scientific tools and procedures that can reduce bias. The imperative of objectivity and independence appears to remain very strong in the evaluation community. For example, Mowles speculates that in using the contingency approach evaluators ‘continue to cleave to what John Dewey [1929]... referred to as a “spectator theory of knowledge”, which assumes separation between the observer and the thing being observed’ (Mowles 2014: 164). Understanding the evaluators’ own value judgements in the role of evaluating an intervention requires making an epistemological shift. It requires moving from a conventional view regarding interventions as ontological realities ‘out there’ subject to evaluation, to a view of interventions as conceptual constructs

(helpfully regarded as systems of interest) and an evaluator as a co-creator of such constructs as part of her/his engagement with the reality.³

In sum, ritualised (mis)understandings in evaluation may involve not only: (i) the piecemeal sectioning of perceived complex situations from complicated or ‘simple’ situations at the outset (that is, failing to recognise that *all* situations involving human activity are complex); but (ii) a possible marginalisation of addressing issues of conflict and equity (again, failing to recognise this as an aspect – to varying degrees – of *all* human activity). In addition, the misunderstanding may involve: (iii) confusing interventions (as well as evaluations of interventions) as ontological realities as against constructivist endeavours; endeavours that speak more to the practice of formative evaluation.

d Evaluating 2: Practice (engaging with multiple perspectives) The formative side of evaluation is related to the actual practice with using tools in an evaluand context and thus impacting on, and generating value with, the stakeholders associated with the evaluand.

In *Aid on the Edge of Chaos* (Ramalingam 2013) a typology of four sets of challenges is offered to those working with complex situations – systemic, behavioural, relational and dynamic. Ramalingam then identifies and draws on four designated families of ‘complex systems approaches’ which can ‘fit’ the designated challenges respectively; ‘systems approaches, behavioural approaches, network methods and dynamic analysis techniques’. This follows a long tradition in the field of management consultancy; working with typologies of situations and then designating methodologies in alignment with the situation-type (*cf.* Ghoshal and Nohria 1993). In effect the process involves not just ‘taming’ situations (e.g. simple, complicated, complex), but taming methods and tools to fit the context as well. In the systems tradition, a similar process gained considerable currency with Total Systems Intervention (TSI) (Jackson and Flood 1991) in allocating particular systems approaches to particular situations. The mapping was done using the Burrell and Morgan (1979) typology of four social science paradigms – objectivist ‘functionalism’ and

‘radical structuralism’, and subjectivist ‘interpretivism’ and ‘radical humanism’. So system dynamics, for example, was assigned to the paradigm of functionalism, whereas soft systems methodology was assigned to the paradigm of interpretivism.

The similarity with the four groups typologised in *Aid on the Edge of Chaos* is perhaps significant, albeit each of the four approaches suggested by Ramalingam relate to large families of methods, rather than any single best practice. Whilst the latter idea is nicely critiqued within international development in terms of ‘an epidemic of bestpracticitis’ (Ramalingam 2013), there remains a pervasive notion of ‘best fit’ – expressed also in the working title of a draft Overseas Development Institute (ODI) paper – ‘From Best Practice to Best Fit’ (Ramalingam, Laric and Primrose 2014).

The problem here remains one signalled by critiques of TSI (*cf.* Ulrich 2003; Reynolds and Holwell 2010; Reynolds 2011). Systems approaches – or indeed any tools, methods, approaches – do not have a pre-determined fixed use-value, outside of the context of use and the individual(s) doing the valuing or acting as users. To assign such ‘objective’ value of tools decoupled from the valuer involves reifying the tools; fetishising the methods. To coin a term used by one popular cynic of command and control approaches used in the public sector, might there be slippage within the evaluation industry towards professionals becoming ‘toolheads?’ (Seddon 2003). Despite warnings in the draft ODI paper of succumbing to a ‘toolkit temptation’ which may lead to unhelpful ‘complexity silver bullets’, there remains a risk in getting fixated on and commoditising complex systems approaches in a ritualised practice of promoting an ever bludgeoning chest of evaluation tools.

There is something here that speaks to a question from Duncan Green ‘... is the whole point of complex systems that you can’t have standard approaches, only connected, agile people able to respond and improvise?’ (Green 2014). These ideas of agility and better practice amongst evaluators is perhaps best captured by a metaphor from the French anthropologist

Levi-Stauss in the 1960s; the art of a *bricoleur*. As Patton explains, this is a travelling nineteenth-century rural craftsman in France skilled in using and adapting tools at hand, but with, as well, a sense of inventiveness and creativity for purposeful intervention – the art of *bricolage* (Patton 2010: 264–304).⁴

e Commissioning (terms of reference) evaluations

Sponsors not unreasonably require some guarantee of rigour from evaluators being commissioned. The terms of reference in commissioning evaluations signal the need for some degree of assurance and trustworthiness. Evidence-based, results-orientated evaluations typically signal guarantors of objectivity – some sense of reliability based on replicability and bias-free recommendations. The popularity of, and investment in, experimental approaches such as RCTs where such assurances are paramount is testament to the power of such guarantors in evaluation. An underlying critique of RCTs from a contingency approach perspective is that they are inappropriately applied to complex situations where the variables generating the effects are often unknown and changeable, and hence difficult to ‘control’. Hence any results generated may not always be reliable.

So what guarantors if any might be offered in complex situations? Glenda Eoyang, the founding Executive Director of the Human Systems Dynamics Institute and a long-time practitioner in the use of complex adaptive systems for evaluation, commented: ‘The only thing you can be certain of is uncertainty, but you can’t even be sure how much of that you have...’. Glenda went on to comment: ‘On the other hand, there’s no escape into constructivist angst and unknowing because there is a world, and we act, and it changes, so there is some grounding for meaningful conceptual models and options for action’ (pers. comm. 2014). Ultimately, of course, there can be no absolute guarantor of certainty with interventions in situations that are inherently uncertain, but there remains a need for commissioners of evaluations to require some sense of trustworthiness.

Practicing evaluators may like to reflect on a need for two sets of co-guarantors in addition to traditional scientific (disciplinary) based

criterion of ‘objectivity’, reliability and replicability. Firstly, there are (more interdisciplinary) co-guarantors of ‘complementarity’ – a communicative competence in facilitating meaningful conversation and exchanges between and amongst stakeholders in the context of an evaluation (process). The guarantor here is one of generating ‘resonance’ with others. These stakeholders might involve other relevant sources of expertise from different disciplines, but critically also including the lay expertise of those directly involved in, and affected by, the intervention being evaluated. A trap in ritualised practice of supporting resonance is in proposing fail-safe techniques that have been ‘tried and tested’ in some contexts and recommended for perceived similar contexts. Unfortunately, the space to experiment – a culture of safe-fail (Stravinsky 2000) – is perhaps an underrated guarantor sparingly afforded to evaluators.

Secondly, there are (more transdisciplinary) co-guarantors of what might be called ‘social and ecological responsibility’ (Reynolds 2008) – a reflective competence in ‘relevance’; not only in addressing the *internal* coherence (interrelationships, multiple perspectives and boundary judgements) of the intervention in dealing with complicatedness, complexity and conflict, but also the *external* legitimacy of the intervention as a whole, in terms of its wider accountability to social and ecological well-being (equity and sustainability).

‘False’ co-guarantors or assurances may involve either incompetence or incapability in the use of any one set, or the unreflective privileging of one set of co-guarantors over another (Reynolds 2001).

- f Learning through evaluating** For commissioners, the immediate need from evaluations is to learn whether and how any particular intervention is working (measures of efficacy and efficiency); learning not only on whether the intervention is serving its purpose but whether the purpose being served is the most appropriate purpose (measure of effectiveness). The learning provided here is in part single-loop – learning to do the thing right – but also double-loop – learning to do the right thing – (see Hummelbrunner, this *IDS Bulletin*).

Evaluators from a tradition using the contingency approach are prone also to advise commissioners in learning not to see *all* situations as being tame but rather to view *some* situations as complex or wicked and requiring different approaches. *Developmental evaluation* (Patton 2010) and *Aid on the Edge of Chaos* (Ramalingam 2013), as well as contributors to the *Evaluating the Complex* compilation (Forss *et al.* 2011), are well endowed with many good case studies illustrating the need for double-loop learning in complex situations.

However, choosing case studies that from the outset are clearly complex involving many viewpoints and conflicts may lend a false sense of security to commissioners that other evaluands (sponsored interventions) are actually ‘simple’ or at most merely ‘complicated’. Perhaps a more significant learning point might come from drawing on case studies that may superficially appear to be merely simple or complicated but actually reveal complexity and conflict in the course of an evaluation. Commissioners might then learn that *all* evaluands are actually complex to varying degrees. This relates to a possible missing third loop of learning regarding power relations (see again Hummelbrunner, this *IDS Bulletin*).

Triple-loop learning questions how ‘rightness’ is circumscribed by ‘mightiness’ and vice versa (*ibid.*). Evaluators may like to consider how the contingency approach has a powerful hold on practice. Arguably, it is an example of relatively unquestioned sense (of ‘rightness’) amongst evaluators dominating (the mightiness of) decision-making amongst commissioners. The contingency approach provides a powerful argument for having multiple methods, each with ascribed fixed functions and values. It prompts the ritualised practice of having an ever increasing ‘armoury’ of methods. It prompts decision-makers (commissioners of interventions, for example) to support an existing toolkit *either* with an ‘overhaul’ of new tools, *or* expanded incrementally with a ‘broader menu of tools’.

4 Discussion: breaking the triangle and abstracting a more purposeful system of evaluation

Whilst there is much to be commended in all three publications using complexity ideas for

evaluating interventions (Patton 2010; Forss *et al.* 2011; and Ramalingam 2013), there remains some doubt as to whether one unholy trinity – one iron triangle – might inadvertently be substituted by another. The first, being implicitly critiqued by all three publications, involves a dynamic depicting: (i) situations regarded as comprising factors related only by linear causality (a judgement of ‘fact’); with (ii) a single viewpoint (value judgement) dominated by a mechanistic perspective of the reality (often caricatured by LFA); and (iii) circumscribed by an associated mindset of command and control (boundary judgement).

The alternative contingency approach proffered by the authors of the above-mentioned publications potentially involves another unholy trinity depicting: (i) situations from the outset as being (‘factually’) either simple or complicated (tame), or complex (wicked), and so with measures of success being ‘fixed’ according to their initial typological demarcation; with (ii) multiple viewpoints each aligned (in terms of ‘value’) with a particular tool, method, methodology or approach; and (iii) associated with a mindset of deceptively benign holism in covering all situation ‘types’, and/or benign pluralism (inclusive of ever increasing tools and methods).

A more purposeful system might have an alternative triadic confluence of the six factors: (a) auditing based on seeing all situations as comprising a mixture of complicated (interrelationships), complex (multiple perspectives), and conflictual (judgements of boundary); (b) planning based on purposeful systems design where measures of success are iterative – flexible and open to change; (c) evaluating (understanding) where values of utility are complemented with intrinsic personal values and political critical values; (d) evaluating (practice) where values are regarded as not fixed but developmental involving the evaluator and other stakeholders in a concerted effort of developing value in existing practices; (e) commissioning based on a wider set of criteria of rigour involving not just reliability and replicability of tools, but the resonance and relevance of tools used in an appropriate critical space of experimentation; and (f) learning based not just on received wisdom but social learning involving all stakeholders in a continual cycling

of single-, double- and triple-loop learning – inviting not just technical, but ethical as well as political insight.

There is evidence of shifts towards a more purposeful evaluation thinking in practice. With respect to (b) ‘planning’, the draft ODI paper (Ramalingam *et al.* 2014) signals the importance of shifting from ‘fixed’ measures of success to measures that are open to revision and change – ‘looking for multiple intervention points and moving between them dynamically... working through group interaction and iteration rather than ‘back office’ designs... [and] concentrating on flexibility rather than predictability of solutions’ (p.11). A key challenge is how these principles might be translated into actual measures to inform a more purposeful intervention. How might they challenge the ritualised practice of fixing measures of success (for intervention managers and evaluators), irrespective of the inevitable changes occurring in situations of the evaluand? Similarly, with respect to (d) ‘evaluating practice’, three other ‘principles’ from the draft ODI paper illustrate well the type of practice required: ‘... accommodating multiple alternative perspectives rather than specific best practices... generating ownership of problem formulation and transparency through participation of stakeholders... [and] developing systematic visual representations of the problem space that enable group-based exploration of the solution space’ (*ibid.*). Also, there is considerable variation amongst practitioners using the contingency approach which does not neatly fit the caricature described above. Many practitioners, for example, use a wide set of co-guarantor attributes – item (e) – over and above assurances of reliable tools. Many also focus on learning – item (f) – in dealing very much with issues of power relations associated with impact evaluations.

A significant part of the Mowles (2014) paper is given to celebrating examples where practitioners in the evaluation field are able to *adapt* ideas and methods using principles of systems thinking and complexity to their own practice. Amongst practitioners cited by Mowles in this endeavour are Allen (1988), Callaghan (2008), Sanderson (2009), Westhorp (2012) and the more contemporary works of Stacey (2012). Other practitioner evaluators whose works are cited by Mowles in exemplifying the contingency approach are contributing authors to these two issues of the

IDS Bulletin (e.g. Hummelbrunner, Rogers, Williams), but their actual practice is much more creative, versatile and adaptive in the use of complexity and systems tools than perhaps allowed for in the main focus of critique. Hummelbrunner (2010), for example, is also cited by Mowles in the same paper as a demonstration of how even a supposedly rigid tool-set such as the Logframe is able to be cultivated for better practice using, in his case, ideas from soft systems thinking. One of the most promising developments of using ideas of complex adaptive systems without presuming a contingency approach are ideas of ‘adaptive action’ (Eoyang and Holladay 2013) based on earlier innovative use of CAS for organisational development (Eoyang and Berkas 1998). Adaptive action resonates significantly with the wider use of systems ideas in our own approach of systems thinking in practice.

Table 2 illustrates these more progressive features in terms of what we might call an evaluation–adaptive complex, in comparison with an evaluation–industrial complex.

5 Conclusion

The identification of three core ideas or concepts associated with systems thinking (*cf.* Williams 2013 and Williams, this *IDS Bulletin*) – interrelationships, perspectives and boundaries – marks a significant shift in abstracting out essential features in order to make systems thinking more resonant with other practices. The three domains correspond to the three issues dominating international development in the last decade (see introduction to *IDS Bulletin* 45.6, 2014). First, there is a growing concern regarding impact evaluation from the policy world; a concern for more evidence-based, results-orientated or outcomes-orientated evaluations, raising issues of rigour in evaluation. Second, there is a concern amongst evaluators with the complexity of interventions driven by multiple agencies and multiple perspectives; concerns that have prompted attention to participatory and multi-method approaches. Third, there is a concern amongst commissioners and evaluators regarding professional responsibility and independence amongst evaluators; a concern leading to an associated need to justify particular approaches.

From a systems thinking in practice perspective, a more general source of anxiety is that such

concerns may themselves become systemically stabilised through ritualised discourse and practice dominated currently in the evaluation field by an omnipresent contingency approach.

The approach offers an alternative degree of rigour towards evaluation, enabling evaluators to adopt particular tools or tool-sets and ideas for particular situations. But the approach arguably has a risk towards stifling the evaluation process by fetishising methods and tools and distracting attention away from craft skills amongst evaluation practitioners and the realpolitik of any evaluation process.

The systems tool used to make this analysis – the systemic triangulator – is one based on ideas of systemic triangulation. As with any tool or model or system – it is an abstraction from reality, and as such is neither right nor wrong in terms of being factually correct. For example, the three

triangular indices of the evaluation–industrial complex (Figure 2) depicting roles of evaluand managers, evaluators and commissioners, may in some contexts actually be embodied in one person. The value of any model ought rather to be measured primarily in terms of its usefulness and adaptability for inquiry in different contexts by different users.

In particular, practising evaluators may wish to reflect and develop value on the use of systemic triangulation for diagnosing possible instances of systemic failure and the role of evaluation in such failure. The article provides a tentative outline model for an evaluation–industrial complex to help with such a diagnosis. To what extent might this be helpful in diagnosing existing evaluation practices, but also to what extent might it help through adaptation in cultivating value towards a more purposeful model of an evaluation–adaptive complex?

Notes

* In addition to the helpful feedback from reviewers on an earlier draft, particular appreciation is extended to Barbara Befani, one of the co-editors of these two special issues of the *IDS Bulletin*. Barbara's insightful suggestions and prompts have significantly enriched the final article. The author remains solely responsible for the final content and any errors.

- 1 See www.collinsdictionary.com/dictionary/english/bourbonism#examples_box.
- 2 That is, the primary intervention – project or programme or policy – of the evaluand (a) and

- (b), rather than the secondary intervention of an evaluation (of the primary intervention) represented by (c) through to (f).
- 3 Callaghan (2008) and Sanderson (2000, 2009) are cited by Mowles (2014: 164) as examples of more reflective thinking of evaluators' roles using systems and complexity thinking.
- 4 Notwithstanding a chapter given to this progressive idea of an evaluator, Patton's book focuses on evaluators as being able to accommodate ever more tools and methods from the tradition of complexity rather than being adaptive and inventive with their existing tool-sets.

References

- Ackoff, R. (1974) *Redesigning the Future: A Systems Approach to Societal Problems*, New York NY: John Wiley & Sons
- Ackoff, R.L. and Emery, F.E. (1972) *On Purposeful Systems*, London: Tavistock
- Burrell, G. and Morgan, G. (1979) *Sociological Paradigms and Organisational Analysis*, London: Heinemann
- Eoyang, G.H. and Berkas, T. (1998) 'Evaluation in a Complex Adaptive System', in M. Lissack and H. Gunz (eds), *Managing Complexity in Organizations*, Westport CT: Quorum Books
- Eoyang, G.H. and Holladay, R. (2013) *Adaptive Action: Leveraging Uncertainty in Your Organization*, Redwood City CA: Stanford University Press
- Flood, R.L. and Carson, E.R. (1993) *Dealing with Complexity: An Introduction to the Theory and Application of Systems Science*, New York NY: Plenum
- Forss, K.; Marra, M. and Schwartz, R. (eds) (2011) *Evaluating the Complex: Attribution, Contribution and Beyond*, Comparative Policy Evaluation 18, New Brunswick: Transaction Publishers
- Ghoshal, S. and Nohria, N. (1993) 'Horses for Courses: Organizational Forms for Multinational Corporations', *Sloan Management Review* 34: 23–35
- Glouberman, S. and Zimmerman, B. (2002) *Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like?*, Commission on the Future of Health Care in

- Canada Discussion Paper 8, www.investisseurautonome.info/PDF-Downloads/COMMENT-INVESTIR-RENDEMENT-INDEX/doc.1590-%20Zimmerman%20Glouberman%202002%20Health_Care_Commission_DP8.pdf (accessed 20 October 2014)
- Green, D. (2014) 'Can Complex Systems Thinking Provide Useful Tools for Aid Workers? Draft Paper on Some DFID Pilot Experiments, for your Comments', *From Poverty to Power Blog*, 7 May, <http://oxfamblogs.org/fp2p/can-complex-systems-thinking-provide-useful-tools-for-aid-workers-draft-paper-on-some-dfid-pilot-experiments-for-your-comments/> (accessed 10 June 2014)
- Hummelbrunner, R. (2011) 'Systems Thinking and Evaluation', *Evaluation* 17.4: 395–403
- Hummelbrunner, R. (2010) 'Beyond Logframe: Critique, Variations and Alternatives', in N. Fujita (ed.), *Beyond Logframe: Using Systems Concepts in Evaluation*, Issues and Prospects for Evaluation of International Development, Series 4, Tokyo: Ministry of Foreign Affairs of Japan
- Hummelbrunner, R. and Reynolds, M. (2013) 'Systems Thinking, Learning and Values in Evaluation', *Evaluation Connections: Newsletter of the European Evaluation Society*, June: 9–10
- Jackson, M.C. and Flood, R.L. (1991) *Creative Problem Solving: Total Systems Intervention*, Chichester: John Wiley & Sons
- Korzybski, A. (1995) *Science and Sanity: A Non-Aristotelian System and its Necessity for Rigour in Mathematics and Physics*, 5th ed. (1st ed. 1933), Englewood NJ: International Non-Aristotelian Library
- Lachica, R. (2013) 'Systems Thinking Usage Survey 2013 Results', *Fuzzy Blog*, 13 March, <http://fuzzyblog.blogspot.no/2013/03/systems-thinking-usage-survey-2013.html> (accessed 7 June 2014)
- Levin, K.; Cashore, B.; Bernstein, S. and Auld, G. (2012) 'Overcoming the Tragedy of Super Wicked Problems: Constraining Our Future Selves to Ameliorate Global Climate Change', *Policy Sciences* 45.2: 123–52
- Mowles, C. (2014) 'Complex, but Not Quite Complex Enough: The Turn to the Complexity Sciences in Evaluation Scholarship', *Evaluation* 20.2: 160–75
- Patton, M.Q. (2010) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York NY: Guilford Press
- Pulitzer, R. and Grasty, C.H. (1919) 'Forces at War in Peace Conclave', *The New York Times*, 18 January
- Ramalingam, B. (2013) *Aid on the Edge of Chaos*, Oxford: Oxford University Press
- Ramalingam, B.; Laric, M. and Primrose, J. (2014) 'From Best Practice to Best Fit: Understanding and Navigating Wicked Problems in International Development', ODI Working Paper May 2014, presented in draft form for comments and feedback, http://oxfamblogs.org/fp2p/wp-content/uploads/2014/05/BestPracticetoBestFitWorkingPaper_DraftforComments_May2014.pdf (accessed 9 June 2014)
- Reynolds, M. (2014) 'Systemic Failure in Macroeconomic Modelling', *International Journal of Applied Systemic Studies* 5.4
- Reynolds, M. (2013) 'Managing Systemic Risk Using Systems Thinking in Practice', *AWERProcedia Advances in Applied Sciences* 1: 217–24
- Reynolds, M. (2011) 'Critical Thinking and Systems Thinking', in C.P. Horvath and J.M. Forte (eds), *Critical Thinking*, New York NY: Nova Science Publishers
- Reynolds, M. (2008) 'Reframing Expert Support for Development Management', *Journal of International Development* 20.6: 768–82
- Reynolds, M. (2001) 'Co-guarantor Attributes: A Systemic Approach to Evaluating Expert Support', 8th European Conference on Information Technology Evaluation, Oriol College, Oxford, 17–18 September
- Reynolds, M. and Holwell, S. (2010) 'Introducing Systems Approaches', in M. Reynolds and S. Holwell (eds), *Systems Approaches to Managing Change*, London: Springer
- Rittel, H.W. and Webber, M.M. (1973) 'Planning Problems are Wicked', *Policy* 4: 155–69
- Seddon, J. (2003) *Freedom from Command and Control*, Buckingham: Vanguard Education
- Stacey, R.D. (2001) *Complex Responsive Processes in Organizations: Learning and Knowledge Creation*, London and New York NY: Routledge
- Stacey, R.D. (1992) *Managing the Unknowable: Strategic Boundaries between Order and Chaos in Organizations*, New York NY: John Wiley & Sons
- Stravinsky, I. (2000) 'Writing the Wrongs: Developing a Safe-fail Culture in Conservation', *Conservation Biology* 14.6: 1567–68
- Ulrich, W. (2003) 'Beyond Methodology Choice: Critical Systems Thinking as Critically

- Systemic Discourse', *Journal of the Operational Research Society* 54.4: 325–42
- Ulrich, W. (2000) 'Reflective Practice in the Civil Society: The Contribution of Critically Systemic Thinking', *Reflective Practice* 1.2: 247–68
- Ulrich, W. (1998) *Systems Thinking as if People Mattered: Critical Systems Thinking for Citizens and Managers*, Lincoln School of Management, University of Lincolnshire and Humberside Working Paper 23, Lincoln: University of Lincolnshire and Humberside
- Ulrich, W. (1983) *Critical Heuristics of Social Planning: A New Approach to Practical Philosophy*, Chichester: Haupt, Bern and Stuttgart, Chichester: John Wiley & Sons
- Williams, B. (2013) 'Systemic Thinking in Evaluation: Three Core Concepts', *Evaluation Connections: Newsletter of the European Evaluation Society*, June: 7–8
- Winckler Andersen, O.W. (2014) 'Some Thoughts on Development Evaluation Processes', *IDS Bulletin* 45.6: 77–85