Learning, Systems Concepts and Values in Evaluation: Proposal for an Exploratory Framework to Improve Coherence

Richard Hummelbrunner

Abstract The three core systems concepts – interrelationships, perspectives and boundaries – can be used for framing an impact evaluation (see Williams, this *IDS Bulletin*). But their use also has implications for the type of learning that an impact evaluation is likely to generate. Moreover, they can help to make the value base of evaluations more explicit. This article first outlines a typology for learning and elaborates on the implications for evaluation and the use of systems concepts. Then a similar typology for values is presented, together with their likely correspondence with learning types. These three aspects are usually dealt with separately in evaluation assignments, although they should be viewed together, as they mutually influence each other or can be seen as complementary. To this end a conceptual framework is proposed which permits to explore and reflect on the connections between the three systems concepts with learning and values.[1]

1 A typology of learning

The *Oxford Dictionary*[2] defines learning as 'the acquisition of knowledge or skills through study, experience, or being taught'. There is a range of theories and a corresponding broad array of learning typologies in fields such as educational psychology, neuropsychology or pedagogy, which explain how learning occurs. Over many decades a rather technical understanding of learning has prevailed in these fields, whereby learning consists in transferring knowledge from someone who has it (e.g. a teacher) to someone who doesn't (e.g. a student). Knowledge can be instilled into students like a substance and there are agreed principles for infusing knowledge. Teachers package knowledge and infuse it in line with these principles. Students are passive recipients whose primary task is to memorise and store that knowledge.

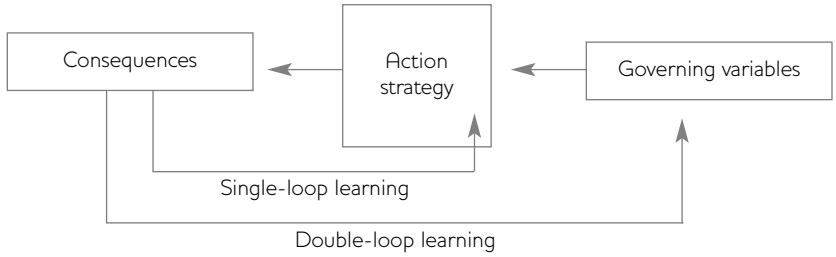
Experience gained with educational practice and new developments in social sciences alike have revealed that this technical model for learning is not appropriate in the case of human beings. According to Gregory Bateson (1972), knowledge in social systems is not passively 'transferred' from a sender to a receiver but actively acquired, namely by selecting information from a wide range of signals ('noise'). Learning takes place by attributing meanings to those signals and transforming some of them into information. The latter is a 'difference, which makes a difference' for the receiver of that information, and learning is the process for marking these differences.

For Bateson, learning also constitutes a process of change with respect to a previous situation. This process can take place at various (learning) levels that are distinguished by the way they deal with trial and error:

- Learning I is the correction of errors by choosing within a specific set of alternatives in a context that remains constant (repeatable).
- Learning II is change in the process of Learning I, which means either a corrective change in the set of alternatives from which choice is made, or a change in how the sequence of experience is punctuated. Learning at this level II assumes that the context in which an activity takes place is

IDS Bulletin Volume 46 Number 1 January 2015
Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

Figure 1 **Single- and double-loop learning**



Source Argyris and Schön (1978).

changing, and it is essentially about recognising the pattern of these context changes and the implications for attributing meaning to an action or behaviour.

- Learning III is change in the process of Learning II, and so forth.[3] Bateson's learning levels represent different orders of abstraction. They can be considered as concentric circles, where each successive level extends beyond the boundary of (and includes) the previous one.

Chris Argyris and Donald Schön (1978) developed these ideas further and applied them to learning that takes place beyond the level of individuals. Their model has been widely used in the organisational development and management literature. The starting point is that people have two mental maps with regard to how to act in a situation. One is the map they claim their behaviour is based on ('espoused theory') and the other is the map they actually apply ('theory-in-use'). Argyris and Schön assert that it is the latter which governs people's actions but that few people are aware of their (often implicit) theories-in-use – and of the difference from their explicit espoused theories. Learning is about reflecting on – and improving – the fit between these two theories of action. To this end they model theories-in-use based on three elements: governing variables, action strategies and actions/consequences. Governing variables are those dimensions (e.g. rules, norms) that people are trying to keep within acceptable limits. Action strategies are used to keep governing variables within the acceptable range: and consequences are what happens as a result of an action.

Learning essentially involves the detection and correction of error understood as a mismatch between these three elements of a theory-in-use. Argyris and Schön distinguish between two types of learning, which are marked by the number of feedback loops involved (see Figure 1).

- *Single-loop learning* occurs 'whenever an error is detected and corrected without questioning or altering the governing variables' (Argyris and Schön 1978). Single-loop learning is equivalent to Bateson's Learning I (as previously outlined). Learning is limited to operational actions that are allowed by the existing norms. Goals, plans or rules are taken for granted and operationalised, but not questioned. Learning is focused on the question of how an existing goal can be reached in the best possible way within existing norms, i.e. by following the rules. It provides short-term solutions to implementation problems but rather deals with symptoms than root causes. The core question is '*Are we doing things right?*'.

  An example often used to explain this concept is the thermostat. It operates in only one mode: when it detects that a room is too cold, it turns on the heater. When it detects that the room is too hot, it turns the heater off. In other words, there is only a limited type of reaction – little or no learning occurs and little or no insight is needed. Single-loop learning is often applied with quality and efficiency problems, for example seeking better ways to manufacture an existing product or deliver an existing service. Organisations operate in a single-loop learning mode when their members establish strategies, policies and procedures and then apply them in a rigid manner by focusing on detecting and correcting deviations from these rules.

- *Double-loop learning* occurs 'when errors are corrected by first examining and altering the

governing variables and then the actions' (Argyris and Schön 1978). Double-loop learning is equivalent to Bateson's Learning II and implies an extra loop reconsidering the generative mechanisms of actions, i.e. the validity of existing norms, rules and assumptions. It works not only at the level of actions, but also of norms – and changing them modifies and expands the scope of available action strategies. Goals, plans or rules are first questioned and then operationalised. Double-loop learning is helpful for making informed decisions when conditions are rapidly changing or uncertain. By addressing the underlying structure of problems (causes and their consequences) it leads to better mid- and long-term solutions. The core question is '*Are we doing the right things?*'

This kind of learning involves critical thinking, 'lateral thinking' or 'thinking outside the box'. Learning in this manner can help people to understand which strategy works better than others to achieve a goal when circumstances have changed. By broadening the scope it also leads to more creativity in developing appropriate solutions. In double-loop learning, members of an organisation reflect on whether norms, rules or assumptions should be changed – and not only on whether deviations have occurred and how to correct them. Double-loop learning is considered to be critical for innovation as well as the continuous success of an organisation or business.

In their later work Argyris and Schön (1996) have investigated why double-loop learning is so rare in organisations. They found that while there is a multitude of espoused theories, theories-in-use fall into one of two models with distinct patterns:

- Model I: The governing variables are achieving defined purpose, winning, and avoiding negative feelings. They also emphasise rationality. The primary action strategy looks to the unilateral control of the environment plus the unilateral protection of self and others. The consequences are defensive routines (often deeply entrenched) that inhibit double-loop learning.

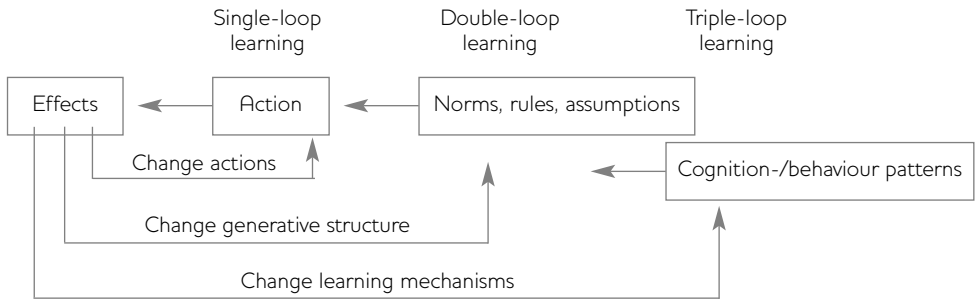
- Model II: The governing values are valid information, informed choice and internal commitment. The primary strategy is shared control as well as participation in design and implementation of action. As a consequence, defensive relationships are minimised and there is a high freedom of choice. This dialogical model fosters double-loop learning.

Argyris and Schön assert that while many people claim to favour Model II, they predominantly apply the defensive routines typical for Model I. This assertion has been validated in subsequent studies by Argyris and his associates, the main reasons being that most people are socialised in Model I behaviour and that organisational rewards are based on values consistent with that model. Therefore, organisations need to make deliberate efforts in order to move people from a Model I to a Model II orientation and practice, for which Argyris and Schön (1996) outline strategies and concrete interventions.

These types of learning address differences with respect to the purpose and extent of learning, and have been expressed in similar ways but with a variety of terms by other authors: lower-level and higher-level (Fiol and Lyles 1985); first-order and second-order (Arthur and Aiman-Smith 2001); incremental and radical (Miner and Mezias 1996); and adaptive and generative learning (Senge 1990). Although these terms stem from different perspectives on organisational learning, a reasonable consensus has been established that they refer to comparable learning processes and outcomes.

Some authors have conceived a further type of organisational learning, for which the most prominent term is *'triple-loop' learning*.[4] Typically, this is described as additional to primary and secondary forms of learning and is often referred to as 'learning to learn', aimed at improving the processes of single or double-loop learning (Tosey, Visser and Saunders 2012). In this case not only actions and norms are considered, but the wider context for learning as well – thus resembling Bateson's Learning III. This wider context can be seen as the paradigm(s) that are at play in a specific learning mechanism. These paradigms can be captured by analysing two kinds of processes: the mental models used to recognise, categorise and interpret sensory input from the environment (cognition), and how these interpretations affect behaviour. By reflecting on cognition and behaviour patterns, changes in a learning mechanism are envisaged.

Figure 2 **Types of learning**



Source Author's own.

By reflecting on the learning mechanisms, existing paradigms are challenged and possibly changed in ways that affect cognition and behaviour patterns. This leads to new ways of identifying and handling problems as well as more sustainable solutions, for example different patterns of recognising and coping with change. The core question is '*How do we establish "rightness"?*'

Flood and Romm (1996) proposed a specific conceptualisation of triple-loop learning that emphasises critical thinking in reflexivity towards processes of learning. Their third loop involves critical reflection on the relations of power that affect cognition and behaviour. Power mandates the learning process and defines the boundaries for single and double-loop learning. It is therefore the key for progress in further learning, that is 'increasing the fullness and deepness of learning about the diversity of issues and dilemmas faced'.

Figure 2 synthesises this typology for learning, highlighting key aspects of the respective definitions and the focus of the loops.

Although each of these types of learning addresses different issues, it is generally assumed that a higher level of learning has greater significance and profundity. Thus the progression from single- to double- and triple-loop learning can be expected to lead to deeper and more sustainable learning.

2 Learning and evaluation

Nowadays, learning is a purpose postulated for most evaluations, be it for improving an intervention while it is still going on (as is the case with formative or developmental evaluations) or for interventions that are to take place at other times and sites or in different contexts (the case of summative evaluations). With respect to the learning purpose, evaluation contracts often specify what is to be learned, who should do the learning or who should be involved in it. But the type of learning that is envisaged within an evaluation is rarely specified and thus remains vague.

Single-loop learning takes place when evaluations function as external regulators (similar to a thermostat): a target is set and a measure (e.g. indicator) established, which allows the evaluator to observe whether implemented actions reach (or are likely to reach) the target. When this measure shows a deviation (due to internal factors or external influences), adaptations are proposed with the intention of correcting actions so the original target is met. But only those corrective options are explored that are compatible with existing values and assumptions. Deviation from targets is *a priori* regarded as negative and in need of corrective actions to put an intervention back 'on track'. Thus evaluation constitutes a negative feedback loop whereby differences from a desired state are counteracted by actions in the opposite direction (i.e. if the value of the measure is too low, then actions are taken to increase it).

Evaluations that function with this single-loop learning mode are capable of finding short-term solutions for specific problems and of changing behaviour or actions needed to that end. But such an approach remains at the level of symptoms and the addressed problems are often dealt with in an isolated manner. Since it does not address the underlying causes of problems,

this approach is rather ineffective (and even counterproductive) in situations of diversity and dynamics.

Double-loop learning occurs when evaluations inquire profoundly about the reasons for deviations from targets, which in turn are connected to programme design and implementation. This includes questioning the definition of goals or targets, deliberating whether expectations on meeting the targets were realistic or what could be done to avoid deviations in the future. In this respect the contribution of existing values, rules and assumptions is deliberately investigated as well and their validity is interrogated. Thus the search for corrective options includes reaching beyond existing implementation frameworks and probing the adoption of new values or assumptions.

Evaluations that operate in a double-loop learning mode are capable of bringing forth profound and sustainable solutions for emerging deviations or problems. They can support ongoing adaptation to developments taking place within an intervention or its operating environment, especially with respect to new or unanticipated situations. They are suited for dealing with complex realities due to multiple and diverse objectives, interests or motivations. And they hold the key to addressing 'wicked problems' (Williams and Van 't Hof 2014).

Last but not least, *triple-loop learning* would take place in situations where evaluations explicitly address the learning mechanisms of an intervention, i.e. reflecting on the observed patterns in recognising or handling problems and in coping with contextual changes. When deliberating on options to change these patterns the underlying power issues should also be addressed, which influence – or inhibit – respective behaviour. Triple-loop learning can also include meta-level discussions about the evaluation with those being evaluated and/or the commissioners of the evaluation. Questions to be addressed in such discussions include how the evaluation was conducted, what was learned in the process and how that learning was produced.

Evaluations (and in particular impact evaluations) are predominantly applying single-loop learning. In order to have more double-loop learning in evaluations, similar defensive routines have to be overcome as in the field of organisational learning. And deliberate efforts are needed to move towards a practice that resembles Model II as outlined by Argyris and Schön (1996): evaluations that envisage a collective learning process, are oriented on dialogue and foster participation. In such cases key stakeholders of an intervention should be involved in the evaluation's design and implementation. And it is their reflective capacity that should primarily be strengthened through the learning processes.

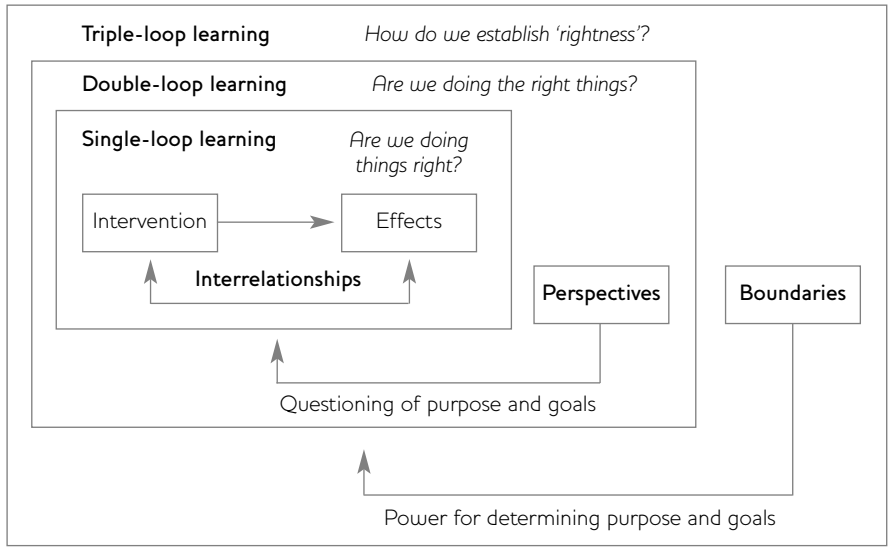
Moreover, treating differences differently plays a key role for achieving double-loop learning in evaluations of social systems (Hummelbrunner 2007). Those systems maintain their balance through continuous renewal – their only constant is change. Therefore differences from original states are inherent to assure their stability. And changes in short-term targets or plans are often necessary for the achievement of long-term objectives. Evaluations of interventions in social systems need to be aware of these characteristics and should not be conceived as an external 'regulator', regarding differences from original plans *a priori* as negative and proposing corrective actions to put an intervention back 'on track'.

Instead, they should value differences from plan (in outputs as well as results and impacts) and also explore their positive aspects. By regarding them as resources and valuable sources of information they can be used to (re)frame observations, or indicate possible alternatives or innovative solutions. Evaluations should also go beyond observing intended effects and look at the entire range of processes or effects, irrespective of whether they are in line with original intentions. Exceptions, discontinuities or unanticipated effects are valuable sources of information about an intervention. They can provide useful clues, for example for relevant internal/external changes, newly emerging challenges, innovative or 'informal' ways of handling situations, which can help to improve implementation.

3 Learning and systems concepts

In his seminal book on organisational learning *The Fifth Discipline*, Peter Senge has argued that key obstacles for learning lie in the different ways of thinking in complex situations. He concluded that higher levels of learning can only be achieved through an improved understanding of complexities as provided by systems thinking (system dynamics or systems archetypes, in his

Figure 3 **Types of learning and systems concepts**



Source Author's own.

case). This is the 'Fifth Discipline' which provides the essential link for the other disciplines of organisational learning: personal mastery, mental models, joint vision and team learning (Senge 1990).

But this utility of a systems approach for learning can be further refined. Based on the typology outlined above, each of the learning loops can be associated with a specific systems concept (interrelationships, perspectives, boundaries – see Williams, this *IDS Bulletin*). Figure 3 illustrates these connections for a generic case of impact evaluation, i.e. assessing the effects of an intervention.

- *Single-loop learning:* The focus is on *interrelationships*, primarily between the intervention and its effects, but also within them (for example between the actions of an intervention or the various effects produced). In case of divergence from original plans, adaptive recommendations are made; for example, modifying a strategy or activities in order to better achieve stated aims and objectives. Significantly, the purpose of the intervention is not questioned.

- *Double-loop learning:* Values and assumptions underpinning an intervention are best reflected if multiple *perspectives* are taken into account in the way these are considered by this systems concept (i.e. stakeholders, stakes and framings). When acknowledging that a situation can be framed in different ways, this also facilitates questioning the purpose and goals of an intervention.

- *Triple-loop learning:* Here the focus is on the boundaries that inevitably have to be drawn in any endeavour, be it an intervention or its evaluation. Reflecting on boundary choices is very helpful (and needed) for critically reflecting on the rules and relations of power that affect behaviour and cognition patterns. This notably involves looking at the power relations that determine the boundaries of an intervention and its evaluation, including the role of evaluation commissioners and evaluators themselves.

This association is one of degree and highlights the respective systems concept that is considered most useful for each learning type. But it should not be understood as a one-to-one mapping, as all of the three systems concepts can be applied in each of the types of learning, depending on the situation. For instance, single-loop learning is not just a matter of looking at interrelationships, and understanding interrelationships can also be useful for double-loop learning, for example for interrogating assumptions about how things work and questioning purposes or goals. Double-loop learning is not just a matter of perspectives, and

examining boundaries can also be useful for reflecting on purposes or goals. But it can hardly be achieved without the systems concept of perspectives, i.e. considering different stakes or framings.

The systems field is methodically quite large, but the various methods and techniques can be aligned with one (or more) of these systems concepts. Therefore, the associations previously outlined can also help to select appropriate methods for evaluations, depending on the envisaged learning type. Here is a brief overview of key contributions offered by the systems field and some exemplary methods (for more details see Williams and Hummelbrunner 2011):

- Interrelationships: The systems field can help in analysing dynamic and non-linear aspects and provides a range of modelling techniques which fall into two broad categories: models that provide insight (e.g. rich picture, causal loop diagrams, social network analysis) or models for prediction (stock-flow diagrams, agent-based modelling).

- Perspectives: The systems field has developed a methodology for looking at situations from different perspectives (soft systems). And it offers techniques for conveying ideas between different stakeholders (e.g. circular dialogue, systemic questioning) and to overcome differences (e.g. dialectical methods of inquiry, solution focus).

- Boundaries: The systems field offers a methodology for assessing the consequences of boundary choices (critical systems heuristic). In addition, some methods contain techniques for addressing specific boundary issues (e.g. viable systems model, container-difference-exchange model).

4 Learning and values

A key role of evaluators is to assign value, thus questions of values and fairness are clearly important in evaluations (Greene 2006). But similar to what has been said above with respect to the types of learning, the values underpinning an evaluation are often vague and are usually not spelled out in an assignment. Thus they largely remain implicit, although some aspects can usually be deducted from the specific approach or methodology applied in an evaluation.

Evaluation is a branch of applied social science and therefore the two main types of values commonly used in that field can be adopted here as well (Kimmel 1988):

- Something has *instrumental value* if and only if it has value as a means to promote some ends. X has value because it is a means to promote some end Y. Money, for instance, has instrumental value, as a means to deliver something else, such as food, clothing, shelter or education. But once it ceases to be a means of getting something else it has no value.

- Something has *intrinsic* (or non-instrumental) *value* if and only if it has value regardless of whether it is also useful as a means to promote some other ends. X has value as an end-in-itself regardless of whether it is a means to promote anything else. This is, for instance, the case with many natural objects, for example a mountain or a lake have a value in itself regardless of other ends they may serve (e.g. resource, pleasure from aesthetic experience).

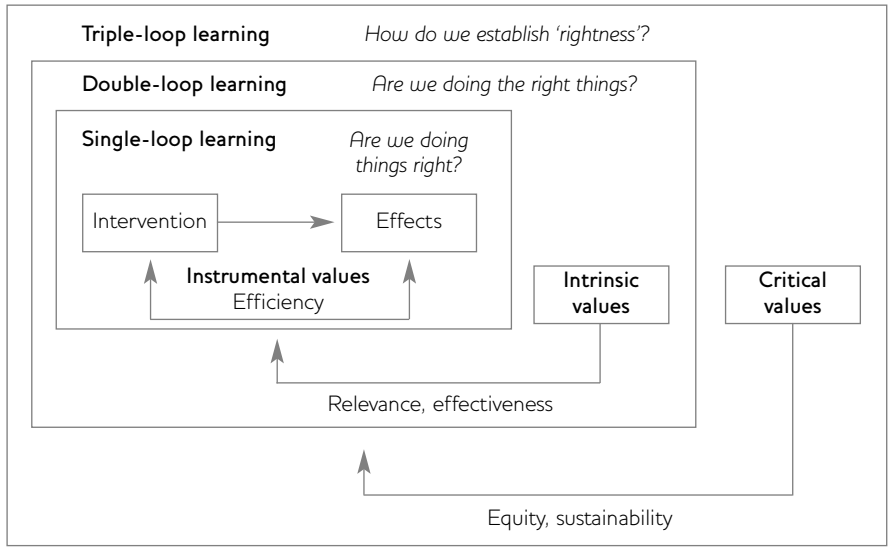
One and the same thing may have instrumental as well as intrinsic value; the two notions can be true of the same object. A person has a value in his or her own right – and this value is taken as the moral foundation of basic human rights. But a person can also have instrumental value, for example an evaluator for those who want to obtain an assessment for their intervention.

In addition, a third set of values might be involved in evaluations that can be labelled *critical values*. These are closely associated with issues of power, expertise and legitimacy. Some evaluation approaches place particular importance on such values: democratic evaluation (House and Howe 2000), for instance, emphasises democratic values like social justice and equity, and critical evaluation theories (e.g. Everitt 1996) propose focusing evaluations around societal critique and structural injustices.

Each of these types of values can be associated with a specific type of learning outlined above. Figure 4 illustrates these connections for a generic case of impact evaluation.

- *Single-loop learning* is based on *instrumental values* embedded in an intervention. These underpin

Figure 4 **Types of learning and values**



Source Author's own.

the intervention logic and can be derived from the respective documents, either explicitly or (probably more often) only implicitly. Instrumental values inform evaluation criteria of 'efficacy' (does it work?) and 'efficiency' (how well does it work using available resources?).

- *Double-loop learning* can surface and reflect the *intrinsic values* underpinning the various framings of an intervention being evaluated. Questioning purpose and underlying assumptions will involve looking beyond the instrumental values of an intervention and take the values of various stakeholders into account, which can be personal, organisational or social values. Intrinsic values inform the evaluation criteria of 'relevance' (why is it important that the intervention works and works well?) and 'effectiveness' (are the right things getting done?).[5]

- *Triple-loop learning* can reveal *critical values*, that is, the values underpinning behaviour and cognition patterns in a particular situation. These are often connected with power relations or issues of expertise and legitimacy – which in turn can be explored when reflecting on boundary choices. Critical values inform evaluation criteria of equity (who benefits from an intervention and why?), as well as aspects of sustainability (are the actors that are needed for sustaining effects adequately involved?).

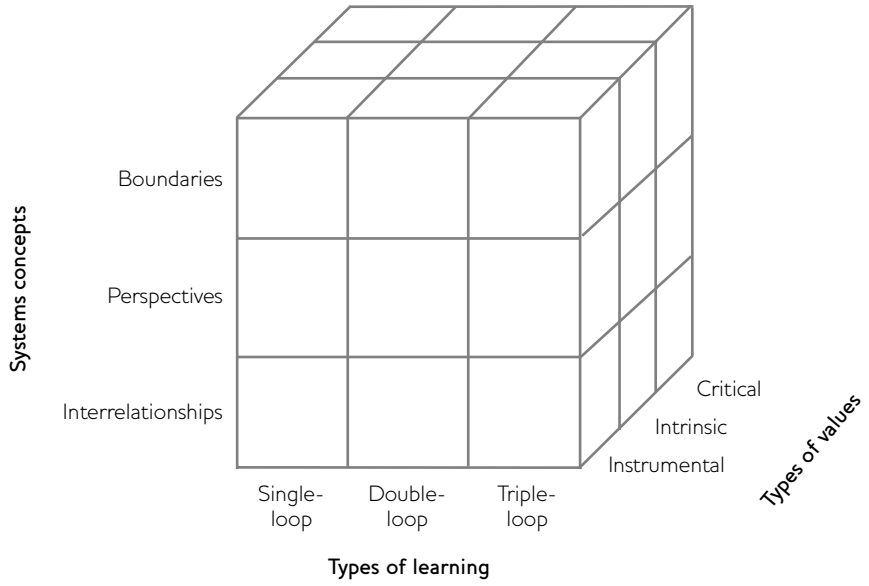
Similar to what has been stated previously, these associations do not imply a one-to-one correspondence between types of values and learning. Values are rarely expressed in a clear-cut manner and thus have to be made more explicit, for example as part of a learning process. Therefore, these associations emphasise which type of value can best be surfaced with a particular learning type. But in an evaluation more than one value type can be applied with a particular type of learning. For instance, double-loop learning not only means to reflect on intrinsic values when examining goals or assumptions, but could also make use of instrumental values (for example in the case of interventions that achieve outputs or outcomes but fail on higher-level results).

5 A framework for assessing coherence

Learning, values and systems concepts can be viewed in an isolated manner – and they are usually dealt with separately in evaluation assignments. But as has been argued in the sections above, they can also be viewed together, as they mutually influence each other or require complementary actions. For instance, the ambition to achieve a specific type of learning has implications for the use of systems concepts (as well as corresponding methods) or on the type of values that can be addressed.

In line with what has been said above, the integration of these concepts cannot be a

Figure 5 An exploratory framework for assessing coherence



Source Author's own.

mechanical exercise that imposes one set of concepts onto another. Fusing together these typologies and their respective implications into a single framework should allow exploration of the space in which these concepts might relate. Therefore the three sets of categories are represented as the axes of a cube: X = types of learning; Y = systems concepts, and Z = types of values (see Figure 5).

The framework can be used in an exploratory manner to interrogate the coherence among the various components of an evaluation assignment: for example, is the type of learning envisaged in line with the evaluation's value base? Can the value base be modified or expanded if a deeper level of learning is envisaged or needed? Which of the systems concepts might be more appropriately applied in making value judgements in an evaluation?

Three situations should illustrate possible applications of this framework:

- The terms of reference postulate that an evaluation should lead to learning in several areas (specified by evaluation questions), and all assessments are to be based on the intervention logic contained in the project document. By making the (instrumental) value base explicit, it becomes clear from the framework that only single-loop learning (doing things right) can be expected from this evaluation and that primarily interrelationships (between intervention and effects) need to be looked into. Double-loop learning could also take place, but needs the consent by the evaluation commissioner for questioning the intervention logic and looking into the assumptions that underlie the programme design and implementation.

- The data gathered during the early stages of an evaluation (or through preceding monitoring activities) shows a rather diverse picture, with many exceptions, contradictions or even puzzles. If the evaluation is to be based on the intervention's instrumental values, the corresponding single-loop learning will not be enough to deal with this situation. Notably, it does not permit exploration of the underlying causes or mechanisms responsible for this diversity in the data (doing the right things). Thus, the evaluator should negotiate with the client as to whether the value base can be changed – and also clarify whether the boundaries (of the intervention or the evaluation) can be questioned if this is not sufficient.

- An intervention appears to be stuck, and due to unforeseen changes the original theory of

action cannot be implemented as foreseen and the implementation context seems rather turbulent and uncertain. Those funding the intervention want to commission an evaluation and are inviting proposals. In this case, triple-loop learning seems to be the best option as it allows a thorough investigation of the boundaries of the intervention. As changes are likely to continue in such a volatile context, the reflective capacity of the implementing agents should be strengthened as well ('learning to learn'), so they can better cope with this situation in the future.

In all of these cases the cube framework can be used to explore what the location of one particular typology (e.g. for learning) implies for the appropriate application of systems concepts and values. Another starting point could be the nature of the situation to be evaluated. During recent years prominent evaluation thinkers (Patton 2010; Rogers 2008) are advocating that evaluation approaches should be matched to the nature of the situation. And they propose a contingency framework that distinguishes between three types of aspects in an intervention:

- Simple: There is high certainty and high agreement between stakeholders about what to do. Cause and effect patterns are clear, predictable and controllable. This is the domain of the 'known', where there is a known right answer (within the current context) and best practice 'recipes' can be confidently recommended.

- Complicated: There is some disagreement about what to do and some uncertainty (due to the involvement of many actors or areas of expertise). The relationships between cause and effect depend on the context, they are neither obvious nor predictable and there are alternative routes to achieve affects. This is the domain of the 'knowable', where good practices can be identified and tested. But answers require careful analysis, coordination and expert knowledge.

- Complex: There is high uncertainty and high disagreement between stakeholders about what to do. The relationships between cause and effect are only evident in retrospect and depend heavily on initial conditions. This is the domain of the 'unknowable', where each situation is unique and previous success provides insufficient clues.

- This contingency framework is gaining increased attention in the evaluation community.[6] When it is applied in an evaluation, it would also have implications for selecting the type of learning or systems concept.

- Single-loop learning is suitable (and most likely sufficient) for the evaluation of simple aspects, namely for analysing deviations from targets or failure of best practices. In complicated aspects double-loop learning can be more appropriate, as inquiring on stakeholder values or assumptions might be required to resolve disagreements and uncertainties. And in the complex domain, triple-loop learning could prove useful, as coping with high disagreement and uncertainties might necessitate addressing prevailing cognition and behaviour patterns as well as power issues.

- The systems concept of interrelationships seems most appropriate for evaluating simple aspects, in particular for understanding cause–effect patterns and the interdependencies between an intervention and its context. The same applies for complicated situations, but more refined methods might be needed for modelling more entangled or non-linear relationships. Engaging with multiple perspectives will also come into play here, for example to analyse or overcome differences with respect to goals, norms or assumptions. And reflecting on boundaries seems indispensable for appropriately evaluating complex situations, because of the utility for dealing with vague or shifting boundaries that are characteristic for this domain.

Again, these are only general implications and should not be understood as one-on-one correspondence. But if the contingency approach is applied in an evaluation and situations are identified accordingly, the cube framework can be used to reflect on suitable learning types, systems concepts and values – here the implications are less obvious and can only be deducted indirectly (via type of learning).

Last but not least, some considerations can be put forth with respect to the utility of this framework for impact evaluation. In recent

years, the challenges in evaluating the impact of complex programmes in international development have led to a rethink of the dominance of quantitative and experimental approaches. A recent study commissioned by the Department for International Development (DFID) explicitly advocates to broaden the range of designs and methods for impact evaluations (Stern *et al.* 2012). A wider range of options could also broaden the scope for the types of learning, values and systems concepts applied in impact evaluations, which in turn could stimulate more thorough reflections on their correspondence with design approaches.

The DFID study recommends that impact evaluation designs are selected in a process aligning evaluation questions with available designs and attributes of the programme to be evaluated. After the design choice is made, the cube framework could be used for exploring suitable types of learning, values and systems concepts.[7] Based on that mentioned previously, the major implications with respect to impact evaluation design approaches appear to be:

- Experimental designs (randomised controlled trials (RCTs), quasi-experiments (QEs)) will only lead to single-loop learning and are based on the instrumental values of an intervention. The same applies for statistical approaches (e.g. statistical modelling, longitudinal studies).

- Participatory approaches can contribute to double- or triple-loop learning and can benefit from applying the systems concept of perspectives. Some of these approaches (e.g. policy dialogue, collaborative action research) address intrinsic (stakeholder) values, while others explicitly advocate the use of critical values (democratic evaluation, empowerment evaluation).

- Theory-based approaches have good potential to contribute to double-loop learning, as they investigate purposes and assumptions. Causal process designs (e.g. contribution analysis, process tracing) are essentially about understanding interrelationships and are predominantly based on instrumental values. Causal mechanism designs (e.g. realist evaluation) embrace multiple perspectives and can be used to bring to the surface intrinsic (stakeholder) values.

- Case-based approaches have similar potential for double-loop learning. Structured approaches (e.g. qualitative comparative analysis (QCA), network analysis) are methods for analysing interrelationships and they are usually based on instrumental values. Interpretative approaches (e.g. ethnography) embrace multiple perspectives and can be used to surface intrinsic (stakeholder) values.

- Synthesis studies can include all of the other design approaches and thus they have the broadest range. They all embrace multiple perspectives and some approaches (meta evaluation, realist synthesis) have good potential for contributing to triple-loop learning.

6 Conclusions

This article applies systems thinking to the evaluation process by outlining typologies for three core aspects – systems concepts, learning and values. It advocates that these three aspects are viewed together more often, as they mutually influence each other or can be regarded as complementary. To this end, a conceptual framework is proposed that can be used to explore a more coherent use of these aspects in evaluation assignments.

This framework can be applied by various evaluation stakeholders (e.g. commissioners, evaluators) for reflecting on the constraints and limitations of an evaluation, as well as for pointing at hidden opportunities that might otherwise be missed. It supports the articulation of the linkages between learning and values, as well as with the systems concepts that can adequately be applied.

Improving the coherence of these three elements can also improve the rigour of evaluation designs, including those for assessing impacts. However, this is not done by privileging certain methods, but by emphasising coherence and congruence across three important aspects. Thus the value base of an evaluation can be made more explicit and congruent with the evaluation mandate. And addressing the appropriate value base (in coherence with the envisaged learning type) can enhance the relevance, validity and credibility of evaluations.

The lack of coherence between the types of learning or values and the evaluation task can

also point at differences between the two mental maps (espoused theories and theories-in-use) of evaluation commissioners and/or evaluators. For example, an evaluation commissioner might claim to want double-loop learning and be open to questioning goals, examining assumptions, etc. but their theory-in-use (i.e. their actual behaviour and the conditions they provide) has a strong leaning towards single-loop learning. When openly confronted with this discrepancy, this might result in defensive (Model I) behaviour. In this case, the use of the exploratory framework can provide the opportunity for achieving a better fit between espoused theories and theories-in-use.

This framework was borne out of theoretical considerations and a deeper investigation into the three aspects, their respective boundaries and potential correspondence. But it is also inspired by the author's practical experience as evaluator, i.e. often being confronted with incoherent or even contradictory demands and expectations in evaluation assignments. In the author's view this challenge is not unique for the design stage, but can emerge any time during the evaluation process, hence the interest in conceiving a heuristic that can be shared with other stakeholders, accompanied by figures and graphs that are simple but not simplistic.

This is still work in progress and at this stage the framework is still tentative and provisional; some connections and their implications might invite further exploration. Therefore, the author would like to invite evaluation practitioners and commissioners alike to test this framework with concrete evaluation assignments, and then share their experience with him, as well as send critical comments or suggestions for improvement.

Notes

1 The ideas in this article were originally presented by the author in a panel session at the workshop 'Impact, Learning and Innovation' at the Institute of Development Studies, Brighton (26–27 March 2013). Furthermore, they were articulated in an article by Hummelbrunner and Reynolds in the June 2013 volume of *Evaluation Connections*, the newsletter of the EES.
2 www.oxforddictionaries.com.
3 Bateson has actually used five levels in his model (from Learning 0 to Learning IV).
4 While triple-loop learning has been inspired by Argyris and Schön, the term does not appear explicitly in their published work.
5 The term 'effectiveness' can have different meanings in different contexts of use; sometimes used interchangeably with efficacy. In this article a clear distinction is made between evaluation criteria of efficacy ('getting things done') and efficiency ('getting things done right') – both of which constitute single-loop learning – and effectiveness ('getting the right things done'). Effectiveness in this latter sense is connected to relevance and invokes double-loop learning.
6 See Reynolds in this *IDS Bulletin* for an assessment of this contingency framework from a critical systems perspective.
7 See Williams in this *IDS Bulletin* on the implications of the systems concepts for impact evaluation.

References

Argyris, C. and Schön, D. (1996) *Organizational Learning II. Theory, Method and Practice*, London: Addison-Wesley

Argyris, C. and Schön, D. (1978) *Organizational Learning*, London: Addison-Wesley

Arthur, J.B. and Aiman-Smith, L. (2001) 'Gainsharing and Organizational Learning: An Analysis of Employee Suggestions Over Time', *Academy of Management Journal* 44.4: 737–54

Bateson, G. (1972) 'The Logical Categories of Learning and Communication', *Steps to an Ecology of Mind*: 279–308

Everitt, A. (1996) 'Developing Critical Evaluation', *Evaluation* 2.2: 173–88

Fiol, C.M. and Lyles, M.A. (1985) 'Organizational Learning', *Academy of Management Review* 10.4: 803–13

Flood, R.L. and Romm, N. (1996) *Diversity Management: Triple Loop Learning*, Chichester: John Wiley & Sons

Greene, Jennifer C. (2006) 'Evaluation, Democracy and Social Change', in I. Shaw; J. Greene and M. Mark (eds), *The SAGE Handbook of Evaluation*, London: Sage Publications

House, E.R. and Howe, K.R. (2000) 'Deliberative Democratic Evaluation', in K.E. Ryan and

L. DeStefano (eds), *Evaluation as a Democratic Process: Promoting Inclusion, Dialogue, and Deliberation. New Directions for Evaluation* 85, San Francisco CA: Jossey-Bass

Hummelbrunner, Richard (2007) 'Systemic Evaluation in the Field of Regional Development', in Bob Williams and Iraj Imam (eds), *Systems Concepts in Evaluation: An Expert Anthology*, Point Reyes CA: American Evaluation Association/EdgePress

Hummelbrunner, R. and Reynolds, M. (2013) 'Systems Thinking, Learning and Values in Evaluation', *Evaluation Connections: Newsletter of the European Evaluation Society*, June

Kimmel, Allen J. (1988) *Ethics and Values in Applied Social Research, Applied Social Research Methods Series* 12, London: Sage

Miner, A.S. and Mezias, S. (1996) 'Ugly Duckling No More: Pasts and Futures of Organizational Learning Research', *Organization Science* 7.1: 88–100

Patton, Michael Q. (2010) *Developmental Evaluation*, New York NY: Guilford Press

Rogers, Patricia (2008) 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions', *Evaluation* 14.1: 29–48

Senge, P. (1990) *The Fifth Discipline: The Art and Practice of the Learning Organization*, London: Century Business, Random

Stern, Elliot; Stame, Nicoletta; Mayne, John; Forss, Kim; Davies, Rick and Befani, Barbara (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper 38, London: Department for International Development

Tosey, P.; Visser M. and Saunders, M.N.K. (2012) 'The Origins and Conceptualizations of "Triple-loop" Learning: A Critical Review', *Management Learning* 43: 291, http://mlq.sagepub.com/content/43/3/291 (accessed 2 October 2014)

Williams, Bob and Hummelbrunner, Richard (2011) *Systems Concepts in Action: A Practitioner's Toolkit*, Redwood City CA: Stanford University Press

Williams, B. and Van 't Hof, S. (2014) *Wicked Solutions : A Systems Approach to Complex Problems*, E-book