The Participatory Evolution Of Alternative Futures

Draft by Dr Rick Davies, 11/5/2020

Abstract

The purpose of this paper is to introduce and provisionally evaluate an innovative web application known as ParEvo. ParEvo is a web-assisted participatory scenario planning process, free to use but copyrighted. The paper begins with an explanation of the conceptual and practical origins of the application design, and its relationship to scenario planning. Two potential uses are outlined, one which is general-purpose and one which is niche specific. The core functioning of the application is then described in some detail. This includes the steps involved in its implementation and the design parameters that can be tuned according to the needs of a ParEvo exercise. Attention is then given to how evaluation and analysis can extract value from a completed ParEvo exercise to achieve both the general and specific purposes. The paper then takes a wider perspective, making some comparisons between the characteristics of ParEvo and other scenario planning approaches. This is done using three different analytical frameworks and a comparator, a scenario planning process that is most similar to ParEvo. ParEvo differs significantly from other scenario planning approaches in three respects: the theory informing the process design, the nature of the search strategy, and the structured and transparent nature of people's participation. The final section looks at the future potential of ParEvo and challenges that need to be addressed, specifically the role of expertise and scalability.

Origins

The immense creativity of biological evolution, spanning more than three billion years, has its basis in what has been called the evolutionary algorithm (Dennett, 1996). Its core features, as originally identified by Darwin, involve the reiteration of variation, selection, and reproduction. More than a hundred years later, Dobzhansky's (1973) claim that "Nothing in biology makes sense except in the light of evolution" is now widely accepted and has even been extended in its ambit. There is now a well-established school of philosophy known as evolutionary epistemology (ref) and there is growing body of machine learning software that embodies various forms of the evolutionary algorithm (ref). The author's PhD thesis on organisational learning built on and extended these ideas (Davies, 1998). One of those interpretations has since survived and proliferated as a now widely used means of impact monitoring in development aid programmes (Davies and Dart, 2005). This paper takes up another variant of methods proposed in that thesis research, now known as ParEvo, which is centred around the implementation of the evolutionary algorithm as a structured social process.

Scenario planning takes place in a wide variety of contexts and a wide variety of methods are available to use (Kobes and Loy, 2020). Arguably, some methods will suit some contexts more than others[1]. The method described in this paper has its origins in a particular economic sector and body of professional practice within that sector. The economic sector encompasses the operations of multilateral, bilateral and non-government organisations in the delivery of humanitarian and development aid to predominantly low-income countries. The particular professional practice is the work of evaluation consultant tasked with the identification of the impact and effectiveness of those aid interventions. With some exceptions, the use of scenario planning methods is not widespread either in this sector or by these professionals. A noticeable exception seems to be the use of participatory scenario planning methods by multilateral and non-government organisations involved in climate change, environment conservation and related economic sustainability initiatives (Nilsson et al, 2019; Planque, et al, 2019; Allington, 2018; Totin et al, 2018; Reinhardt et al, 2018; Oteros-Rozas, 2015;). More recently there have also been some papers looking at the relevance of scenario planning to evaluation (Frith and Tapinos, 2020; Ramirez and Brodhead, 2019; Derbyshire, 2019).

The evaluation of development aid interventions is typically a backwards-looking exercise. The generic purpose is to identify what people were trying to do, how they went about it and what were the consequences, as seen from different stakeholder perspectives. At best, the planners of the development intervention will have articulated what is called a Theory of Change (Vogel, 2012). This is usually a fairly formalised representation of the nature of the interventions planned, and their short, medium and longer term anticipated effects, along with some reference to surrounding contextual factors that may constrain or enable the intervention to work. These representations range from very simple chain models (where activities lead to outputs leads to outcomes leads to impacts) to quite complex network models involving many different forms of feedback (Davies, 2018). The designers of these representations must balance two competing requirements. One is that the representation has sufficient detail to make it evaluable i.e. measurable and testable. The other requirement is that the representation is sufficiently simple to be easily communicated and understood by those funding and implementing the intervention, and by other stakeholders. This is not an easy task. Relatively sophisticated methods such as systems dynamics modelling (Craven 2017) and agent-based modelling (CECAN, 2016) can have a better fit with the complexity of the real world, but their workings are not easily communicable to non-specialists.

In the background there is a more demanding requirement that has been gaining more attention by aid agencies. This is that the Theory of Change must cope with the fact that the world is not just complex but often unpredictable (Buffardi, 2016). In these circumstances, development interventions need to have the capacity to recognise and adapt to changing circumstances. Major international aid donors such as the United Kingdom's DFID and the United States' USAID programs have been funding applied research programs aimed at identifying ways of building this capacity within international aid programs (GLAM, 2020; CECAN, 2020). Some of these research programs have recently explored the relevance of scenario planning (Young et al, 2019).

ParEvo: A new approach to scenario planning

ParEvo is a web -assisted participatory scenario planning process. It was designed by the author in 2019 and then encoded as a web application by Aptivate, a UK software development consultancy firm[2]. While the process has been copyrighted the use of the application is free of charge. Information on the design and use of the application is available on a dedicated website[3]. Eight ParEvo exercises have been completed to date, predominantly as pre-tests of the design and then beta tests of the developed application.

ParEvo has two types of potential uses. The first is relatively niche specific. This is to enable the designers, implementers, and evaluators of development aid programs to think more clearly, flexibly and adaptively about the futures that those programs may encounter. How this can be assisted is discussed further below.

The second purpose is more widely applicable. This is to provide what is in effect a research platform where the process and products of multiple ParEvo scenario planning processes can be documented, compared, and analysed with the aim of accumulating knowledge about to do useful scenario planning exercises. In the process the design of the ParEvo app is also likely to evolve. How this can be enabled is also discussed further below.

Core features

The core of the ParEvo process is a social version of the evolutionary algorithm, as shown in Figure 1 and then explained further below.

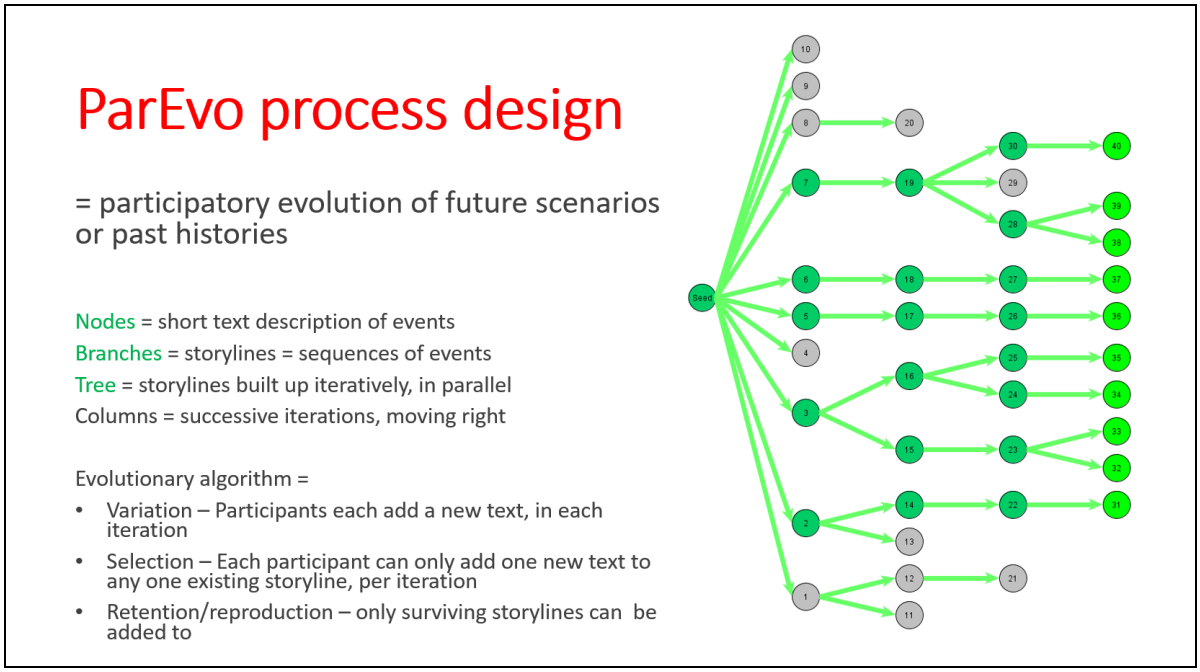

Figure 1: The ParEvo process as a version of the evolutionary algorithm

A ParEvo exercise proceeds by participants making contributions through a series of iterations, each of which is shown as a column in figure 1 above and proceeding from left to right. An exercise begins with a 'seed' text, which is a paragraph describing an initial event, usually provided by the Facilitator of the exercise. The exercise participants, ten in the imagined example in Figure 1 above, are then invited to each submit an additional paragraph describing what they think happens next. Thus, in the second column from the left there are now ten alternative storylines in the making. In the next iteration the same participants review those ten storylines and choose (only) one which they would most like to continue - by adding a third paragraph. In Figure 1 we can see that three of the initial storylines were ignored by all participants and three others were each extended by two participants. In the next iteration the same process was repeated, some developing storylines were ignored, and others continued and some of those were continued were added to by more than one participant. Each storyline is one scenario, a narrative account of what could happen in the future given the original seed paragraph and how it was added to thereafter. Some storylines become extinct, some survive, and some proliferate. This is a barebones description of how the ParEvo process works. How well it works depends on various parameters set by the Facilitator but also on how the participants themselves respond to the opportunities arising in each iteration. It should be noted that participants cannot identify who contributed which paragraph, but the Facilitator can.

The parameters for a ParEvo exercise fall into four broad categories:

- Definition of exercise objectives
- Participants to be involved and their roles (selves, stakeholder representative…)
- Number and duration of iterations for the exercise
- Content of the Facilitator's guidance, provided at the beginning and during the exercise
- Evaluation criteria, to be used by participants at the end of the exercise

The ParEvo application has two interfaces, one which is visible to the participants shown in Figure 2 below, and another which is visible to the Facilitator only. The participants' interface has four sections. On the top right there is the Facilitators guidance to the participants, updated prior to the beginning of each new iteration. On the left is an optional image, associated with the topic of the

exercise (described by the headline). Below the image is a clickable tree structure showing the original seed paragraph at the top and all the subsequent additions to date below. Two different types of storylines are visible in Figure 2. Extinct storylines are shown in grey, and surviving storylines are shown in green. To the right of the tree structure there is a body of text which is the content of the storyline highlighted in light green on the left. Users can access the content of any storyline by clicking on the end node of any storyline in the tree structure. The contents of the user interface can also be viewed by Facilitator-designated 'observers ', but they cannot add, amend, or delete any content.

The ten steps involved in the development implementation and completion of the typical ParEvo exercise, as shown on the ParEvo website[4], can be reduced to the following five:

1. Identification of ParEvo exercise objectives
2. Identification and recruitment of participants
3. Exercise parameters set by Facilitator
4. Initiation of exercise: Reiteration of variation and selective reproduction of storylines
5. Participants evaluate surviving storylines
6. Analysis and use of the results

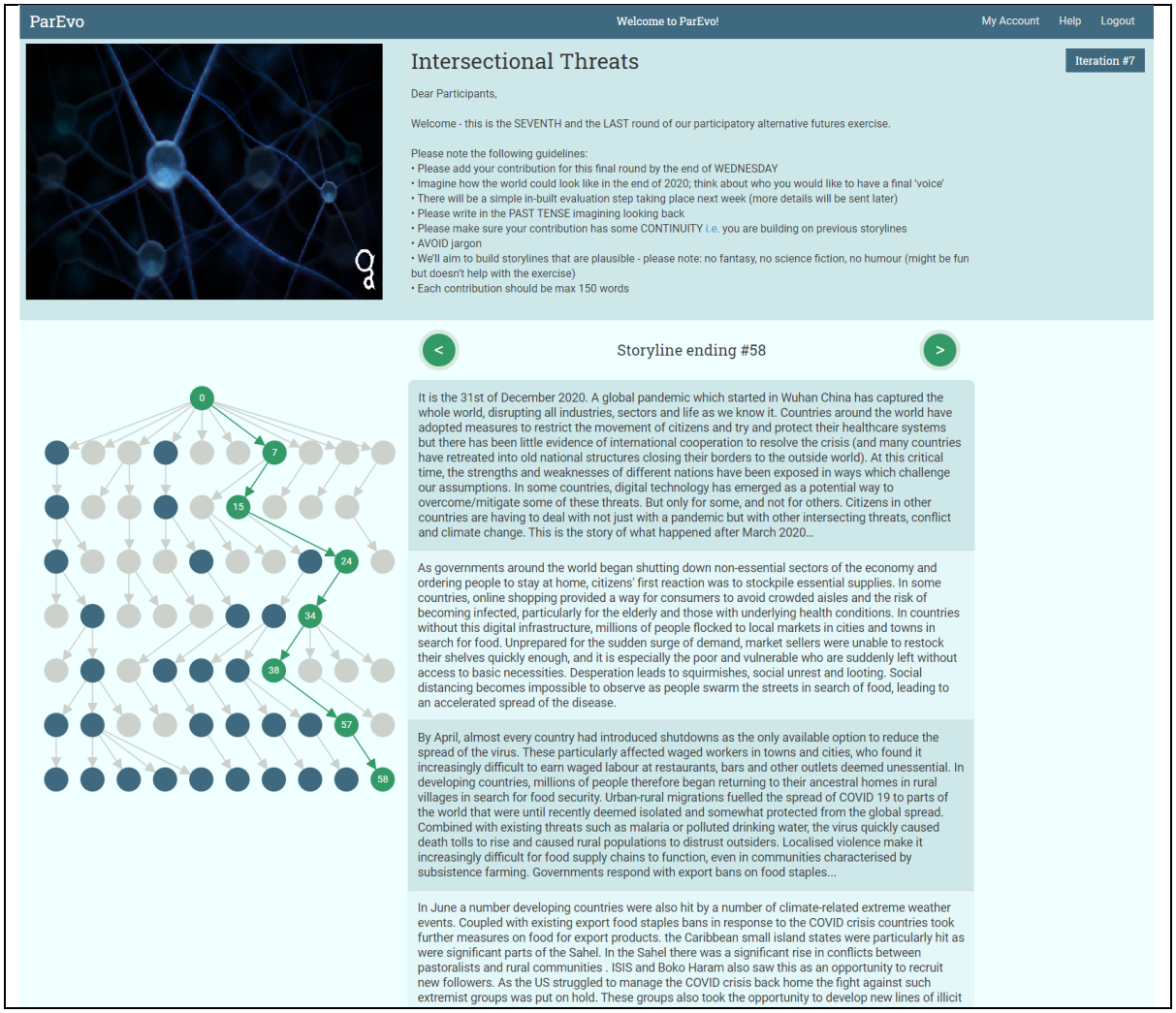

Figure 2: The ParEvo user interface, with an example exercise in progress

Evaluation and Analysis

Participatory Evaluation

Evaluation is the ninth of the ten steps involved in the use of ParEvo. Participants can evaluate the results of a ParEvo exercise in two ways. The first is via an evaluation feature that is built into the app and which is triggered by the Facilitator, typically after the last iteration is completed. At that point, a menu appears inviting participants to identify individual storylines that they think is: (a) Most likely, (b) Least likely, (c) Most desirable, (d) Least desirable. These are the default evaluation criteria. The Facilitator can edit the evaluation criteria presented by this feature, if there are others more relevant to a particular exercise. In the final stage of the ParEvo exercise participants' aggregated evaluation responses for each storyline are displayed on the user interface.

The second evaluation option is for the Facilitator to use a third-party online survey service, such as SurveyMonkey, to ask a wider or detailed set of questions. The primary use of this option to date has been to enable the Facilitator to ask what are called pile sorting or card sorting type questions (Coxon, 1999). Such as: '*Please sort this list of storylines into two piles of any size each, according to what you think is the most significant difference in their content. Then please describe the nature of this difference'*. This more ethnographic form of enquiry enables Facilitators to identify emergent characteristics of the storylines that they may not have thought of or prioritised beforehand (Spradley, 1979).

Analysis by the Facilitator

Five types of data can be downloaded by the Facilitator after a ParEvo exercise is completed:

- A participant x participant matrix, showing who added on to whose paragraphs
- A storyline x participant matrix, showing which participant contributed to which storyline
- A storyline x evaluation criteria matrix, showing the number of times each storyline was identified as having each evaluation criterion
- The full text of each storyline
- A list of participants

This data can and has been used to carry out two types of analyses: (a) the content of the storylines, (b) the structure of people's participation in the exercise. Two pages of the ParEvo website are being used to accumulate knowledge of how this can be done[5].

Content analysis

At present the content analysis page provides information on, and examples of, four methods of analysis: (a) Participatory, using forced-choice questions to identify storylines that are most exceptional on some predefined evaluation criteria – see above, (b) Participatory, using a pile-sorting exercise to identify clusters of storylines with similar content, and to capture what characterises each cluster– see above, (c) Specialist, where the facilitator (for example) codes the contents of each of the contributions to each of the storylines according to themes of prior and emerging interest to themselves or their clients, (d) Machine learning (using a 'topic modelling ' algorithm) to identify clusters of storylines having similar content.

Figure 3 shows an example scatterplot visualisation of participants evaluation judgements regarding seven storylines in one of the pre-test exercises. The position of a storyline reflects the net judgements of the participants e.g. the number of participants who said a storyline was most desirable minus the numbers participant said it was least desirable, and the number of participants

who said a storyline was most likely minus the number who said it was least likely. Storylines on the edge of the plot are those who status is most agreed on, whereas those in the centre of those who status is more disputed, or unclear (not identified as being most or least on either criteria).

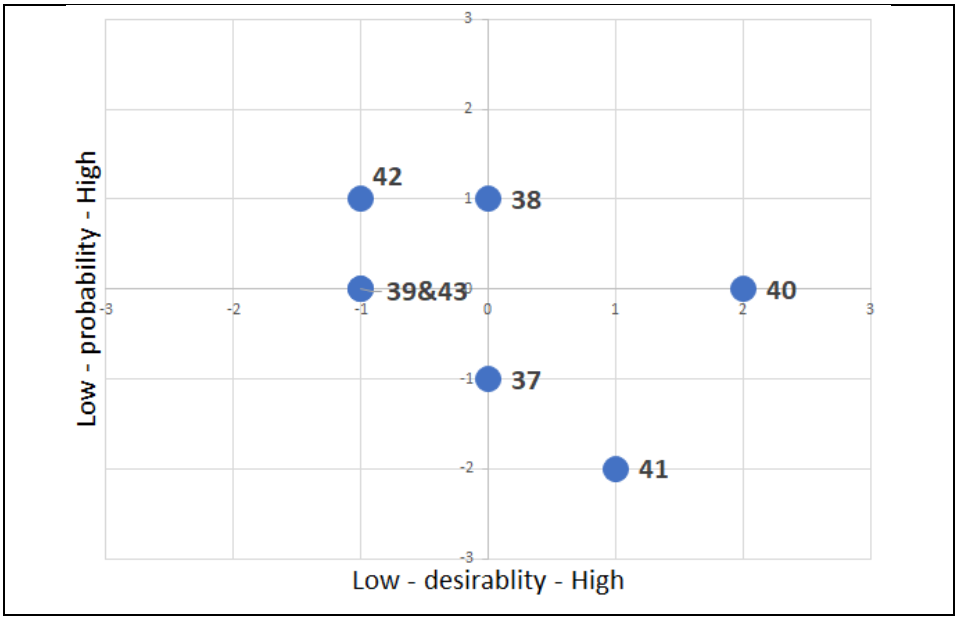

Figure 3: The distribution of participants' evaluation judgements, from a 2019 pre-test exercise. Storylines are shown by number.

Figure 4 is an example network diagram showing the co-occurrence of different coded themes found in ten storylines in a recent beta testing ParEvo exercise. Numbers next to the nodes refer to the number of times that theme was mentioned, numbers on the links between nodes referred to the number of times the two linked themes co-occurred.

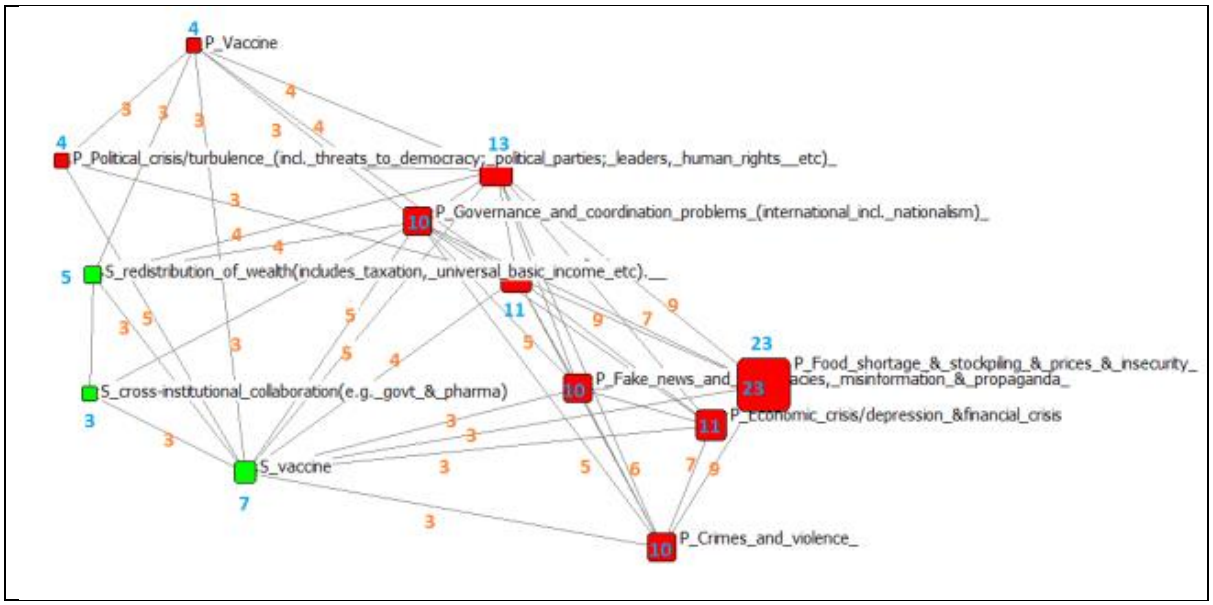

Figure 4: Network diagram showing co-occurrence of selected red "problem" and green "solution" themes found in the contents of a set of storylines in a beta-test ParEvo exercise.

Participation analysis

There are two reasons for paying attention to the nature of people's participation in a ParEvo exercise. The first is an instrumental one, an analysis of how people participate may help identify ways in which the contents of storylines can be improved. The second is a concern about participation for its own sake. An analysis could help achieve objectives relating to improving how people participate, e.g. as a cohesive group versus a set of cliques.

The participation analysis page explains, with examples, how downloaded participation data matrices can be analysed using social network analysis tools, such as Ucinet (Borgatti, et al, 2002). Both the ways in which participants selectively build on their own and other participants contributions to the storylines, and the relationships between different participants' evaluation judgements. The analysis examples currently available also show how measures of diversity, drawn from the field of ecology, can be used to compared participants behaviour, and the content of the storylines themselves, within and across ParEvo exercises. This focus on diversity measures was prompted by a reading of the literature on collective intelligence (i.e. how a group of people can be more effective than the most effective individual in the group) (Yu, et al, 2018) and related research on the effects of diversity on effectiveness (Page, 2017).

Figure 4 below shows two examples of social network diagrams, and associated network metrics, that map the structure of collaboration seen in two pre-test exercises. It should be noted that these structures are essentially emergent rather than planned because the participants were not aware of which contribution was made by whom during exercise. Because of this these structures might be better described as networks of ideas rather than actual social networks.

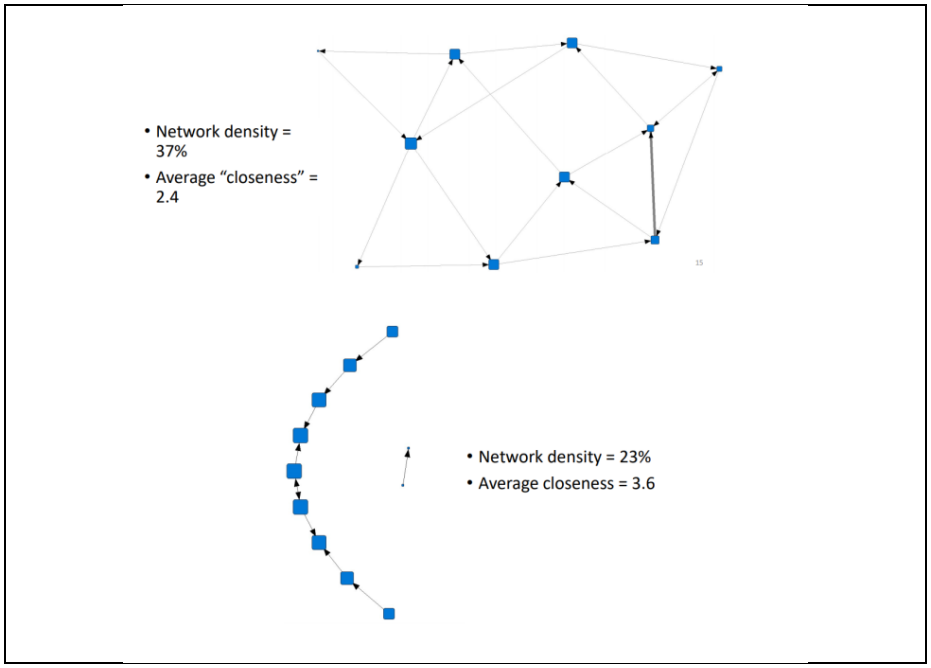

Figure 5: Collaboration networks from two different pre-test ParEvo exercises. Nodes = participants, links = one participant adding a paragraph to that of another participant

Application

It is expected that the contents generated by a ParEvo exercise could be used at three stages of a development project or intervention of some kind: (a) planning, (b) implementation, (c) evaluation[6].

At the planning stage there are two possible uses. One is to help articulate in implementable and evaluable detail the contents of the causal pathways in a Theory of Change, from intervention to intended impact. The other is to help develop a risk management plan that covers two types of scenarios: (a) scenarios where some relative probabilities can be assigned, and (b) scenarios where it is not possible to do so. Both of which can be identified from the participants' evaluation judgements at the end of a ParEvo exercise. This distinction is some correspond with an important distinction often made between risk and uncertainty (Mitchel, 2009).

At the implementation stage the main anticipated use of the identified storylines is to help widen the scope of an organisations 'peripheral vision', and thus to identify the arrival of any foreseen risks and opportunities as early as possible.

At the evaluation stage there are two potential uses. At the macro level, involvement of a range of stakeholders in a ParEvo exercise at the time of an evaluation could help identify important differences interpretations of the key moments in the history of the intervention, which the evaluation team will need to pay attention to. At the more meso level the detailed provided in a individual storyline will provide a useful starting point for process tracing types of analysis aimed at identifying the causal mechanisms responsible for particular outcomes of interest[7].

At present these applications remain as possibilities, yet to be tested because an appropriate ParEvo exercise opportunity has not yet become available. In contrast, the content analysis and participation analysis methods described above have already been applied to completed exercises. Thus, some initial progress has been made towards ParEvo's second overall objective i.e. the development of the research platform that is accumulating knowledge. The details of that knowledge are not discussed in any further detail in this paper because its purpose is more introductory and comparative.

Comparisons with other approaches

Is ParEvo a form of scenario planning?

In the face of allegations of terminological confusion about the nature of scenarios Spaniol and Rowland (2019) have designed a decision tree to enable the identification of what constitutes a scenario (see Figure 6). Working through the set of questions in the decision tree it seems clear that a set of storylines generated by a ParEvo exercise does qualifies as a set of scenarios. One possible area of doubt is the status of 'systematised set'. But some of the example descriptions given for this status do seem to fit a set of ParEvo storylines: "*Numbers of scenarios combine to become a "set of shared and contrasting narratives" (…) or even multiple "sets of narratives about the future" (*...) *While "system" can refer to a network of actors or even "value systems" (*...)".

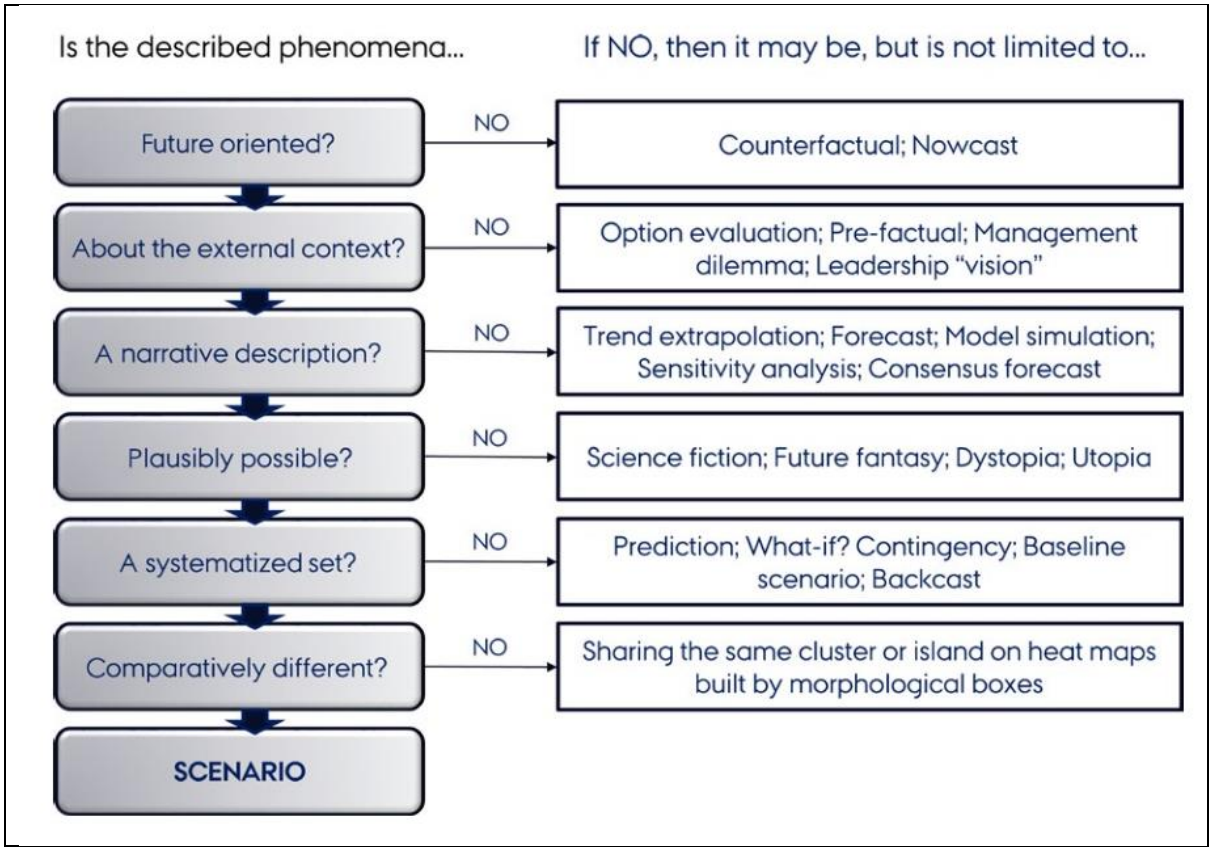

Figure 6: Process for classifying a phenomenon as a scenario in the Intuitive Logics tradition (Spaniol and Roland, 2019)

Typologies of scenario planning approaches

A three-dimensional model

In Bengston's (2019) review of "futures research methods and applications in natural resources" he provides a useful three-dimensional classification system, as shown in Figure 7 below. Within this framework ParEvo can be identified as being: (a) more qualitative than quantitative, (b) more imagination based than evidence-based, and (c) more participatory than expert based. However, there are some important caveats with this description. Firstly, as already suggested above, there are multiple means of undertaking quantitative analyses of both the ParEvo storyline contents and the participatory process. Secondly, when the Facilitator sets the parameters for an exercise one of these can be the extent to which the participants contributions should be about probable rather than just possible developments. Probable developments will require more of an evidence-base. Not necessarily one provided within the individual contributions made by the participant, but more likely one that is thought likely to be available in the background. Thirdly, while ParEvo is pre-eminently a participatory approach, the process can include experts. They can provide initial briefings to participants; they could be included as one type of participant and they can be involved in the final evaluation of the surviving storylines.

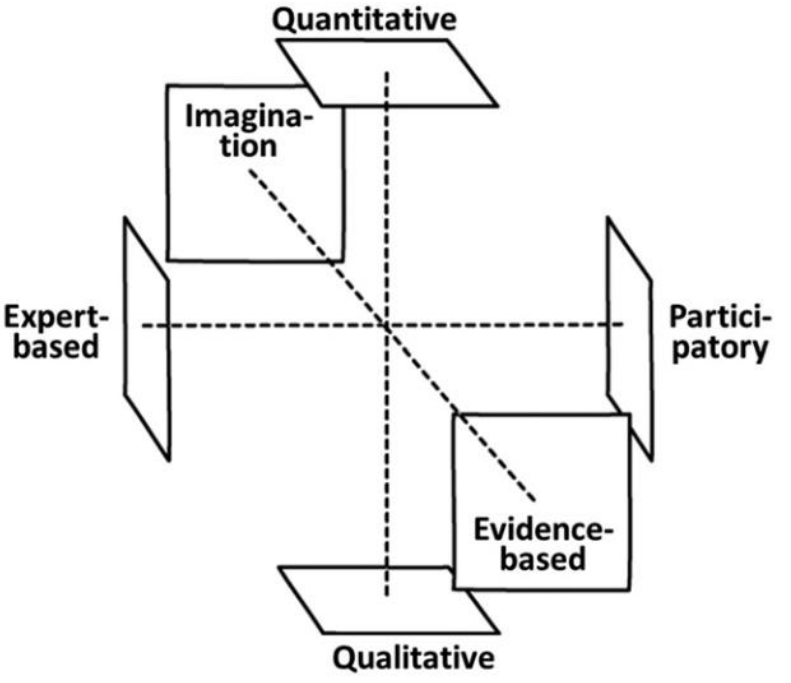

Figure 7: Three dimensions of futures research methods (Bengston, 2019)

Performance distinctions

Seeve (2018) has provided a different set of distinctions, which are more evaluative than descriptive. He describes four different types of criteria for assessing the performance of scenario planning exercise.

- Explorativeness: to enable thinking of the unprecedented, the process must be sufficiently explorative.
- Trustworthiness: to establish trust in the developed scenarios, the process must be transparent and consensus supporting.
- Efficiency: to develop scenarios in a timely and efficient manner, the process must be well structured.
- Accessibility: to ensure that the stakeholders can contribute their best expertise, the process should not push the stakeholders to far from their capacity.

ParEvo performs well on first criteria. Each exercise involves the evolution of multiple storylines. The default setting is that number of storylines equals the number of participants, a parameter that can be changed. This contrasts with the more limited diversity of scenarios produced by methods that used 2x2 matrices to combine different drivers of change. Bengston (2019) describes the latter as characteristic of the Intuitive Logics method, one of the more well-known and long-standing approaches.

On the second criteria ParEvo performs well on transparency but not on the consensus criteria. In contrast to many participatory methods of scenario planning it is possible for the Facilitator to identify the individual contributions of each participant throughout the whole exercise. This can be done without sacrificing the anonymity of contributions, which it is been found is an important

means of maintaining diversity, a key ingredient in any 'wisdom of the crowds ' approach (Yu, et al, 2018).

On the consensus criteria, ParEvo deliberately fails. There is no intention to deliver consensus opinion at the end of an exercise. The very idea appears to contradict the aim of identifying multiple alternative but plausible scenarios. Nevertheless, evaluative judgements are sought on the different scenarios that are developed, and it is possible to identify the degree of agreement between participants on these judgements.

ParEvo performs well on the efficiency criteria. Because it is a web-based process people who are physically distant can nevertheless participate in the same process. The main constraints on the duration of the process are differences in time zone, which means some participants cannot contribute until many hours after others have done so. The time required to read the contributions that have been made in the previous iteration is not substantial. And the time required to make new contributions is also modest. A ParEvo process typically takes place over a number of days with a small input on each of those days. Rather than taking up a whole day or two. In that respect it can be much easier to schedule around people's existing work commitments.

The ParEvo interface has been designed with accessibility and usability very much in mind. Compared to many web applications the interface is quite simple and unsophisticated. If the process does stretch people's capacities it is their imaginations that are being stretched rather than their technical knowledge. Questions could be raised about the capacities of participants to make sense of large numbers of storylines, when choosing which storyline to develop further or when asked to make evaluative judgements at the end of the exercise. This issue is discussed further, under Scalability below.

Inductive versus deductive approaches

Seeve (2018) and others (McBride et al, 2017) have identified another key difference between scenario planning approaches, the difference between the deductive and inductive approaches. Deductive approaches develop scenarios via a general framework by first identifying the most influential and uncertain drivers of future change and then selectively combining what they see as the most important of these drivers in a 2x2 matrix format. Narratives are then developed for each of the four scenarios.

Seeve (2018) has identified the weaknesses of this approach, in terms of the performance criteria already discussed above, as follows: ".. *With respect to the explorativeness aspect, the deductive 2×2 scenario matrix approach is limited in that (a) focusing on the extremes of axes can drive unnecessary polarization in thinking and (b) limiting the number of uncertainty factors at an early stage may pre-emptively restrict the exploration of the future possibility space… With respect to the trustworthiness aspect, the deductive approach can be limited as well. More specifically, reaching consensus on the selection of the two scenario axes may be difficult, which may lead to a lack of trust by the process stakeholders in the scenarios resulting from these axes … For example, van't Klooster and van Asselt (2006) observed controversy in the selection of the scenario axes in the scenario projects they followed. Stakeholders of scenario workshops felt a lack of trust in the selected axes, their critiques including, for example, that 'the scenario axes represented a very classical scheme', and that the selection criteria of the two key uncertainties were not transparent*"

Perhaps contrary to expectations, this critique also seems to be applicable to many participatory scenario planning processes. In the Oteros-Rozas (2015) review of twenty-one participatory

approaches the majority of cases began with the identification of key drivers, which then led to the development of narratives based on selected drivers or combinations of those.

In contrast to deductive approaches, inductive scenario planning methods use specific-to-general techniques for building scenarios. Here "*… scenarios [are] emerging from in depth discussions about individual events or plot elements, around which larger scenario storylines are then developed organically*" (McBride et al, 2017) – a description that easily fits the process of storyline development in a ParEvo exercise. In addition, the evaluation and analysis stages in the ParEvo process enable the subsequent identification of the more abstract and general features of the storylines, including drivers of change and their interconnections (see Figure 4).

Seeve (2018) also make some useful critiques of inductive approaches. "*If the inductive scenario development exercises are carried out in an unstructured manner … the process can become more opaque and dependent on the creativity and imagination of the participants. This unstructured nature of inductive processes can cause greater time and facilitation demands and even risk the success of the scenario exercise.... Thus, inductive scenario development approaches can be inefficient, having limitations with respect to the third aspect (efficiency) of scenario technique selection... Moreover, opaqueness in building scenarios can undermine trust in the developed scenarios, and thus an unstructured inductive process can have limitations with respect to the second aspect (trustworthiness) ...*"

However, while ParEvo is an inductive process it is without some of the weaknesses described above. The time frame can be carefully circumscribed by a facilitator. ParEvo uses a very structured process, and the process is transparent, and can be efficient and trustworthy – as discussed earlier. But it is still true that the process is dependent on the creativity and imagination of the participants.

Place-based versus Internet

It might be expected that a review of this kind would make some comparisons between place-based face-to-face scenario planning processes and Internet-based processes. On the surface, face-to-face processes have the possibility of providing a much richer communication medium for the participants. On the other hand, Internet-based processes have the possibility of greatly extending the number of participants. In reality, these distinctions are more blurred. Video-based participation is becoming more widely used and accessible. Face-to-face processes can also engage large numbers of people, particularly those designed to be participatory (Oteros-Rozas, 2015). Processes like Delphi, which were originally face-to-face based events, have been adapted for use via the Internet (Agustan et al, 2019). This suggests that the key question that should be asked here is how adaptable is the web-based ParEvo process to situations where face-to-face based interactions are needed or preferred.

One pre-test has been carried out, which is suggestive of the possibilities. This was in circa 1996, in the course of the PhD research on organisational learning mentioned earlier. The setting was a classroom of secondary school students in their final year, in Wales. Ten of those students were involved in a ParEvo exercise with essentially the same as that described earlier in this paper. The key difference was that the seed paragraph, and all subsequent paragraphs that were added on, were written down on filing cards and placed on a blackboard, along with chalk lines showing which paragraph with building on which paragraph. Within the space of an hour the process went through four iterations. The topic was a light-hearted one, about the future of one of the students in the class after left school at the end of the year. The process proceeded without difficulty and has been documented[8]. In retrospect, the key challenge faced in this type of exercise is how to preserve the

anonymity of the contributions. In the 1996 pre-test everyone could see who had put out which contribution. Perhaps not surprisingly the subsequent network analysis of how people participated showed some cliques, where three participants repeatedly built on each other's contributions.

The preservation of anonymity in physical settings like this should not be a major problem. Participants contributions could be channelled through the hands of the Facilitator, and then posted all at the same time. Since 1996 the technical options available for the physical display of the emerging storylines have expanded, and should not represent a significant constraint.

The nearest comparator

The scenario planning process most similar to ParEvo is the Futures Wheel (Glenn, 1972), also known as the Consequences Wheel or the Implications Wheel. Figure 8 is a simplified example of a Futures Wheel. Both have a branching tree structure, which develops out from a single origin.

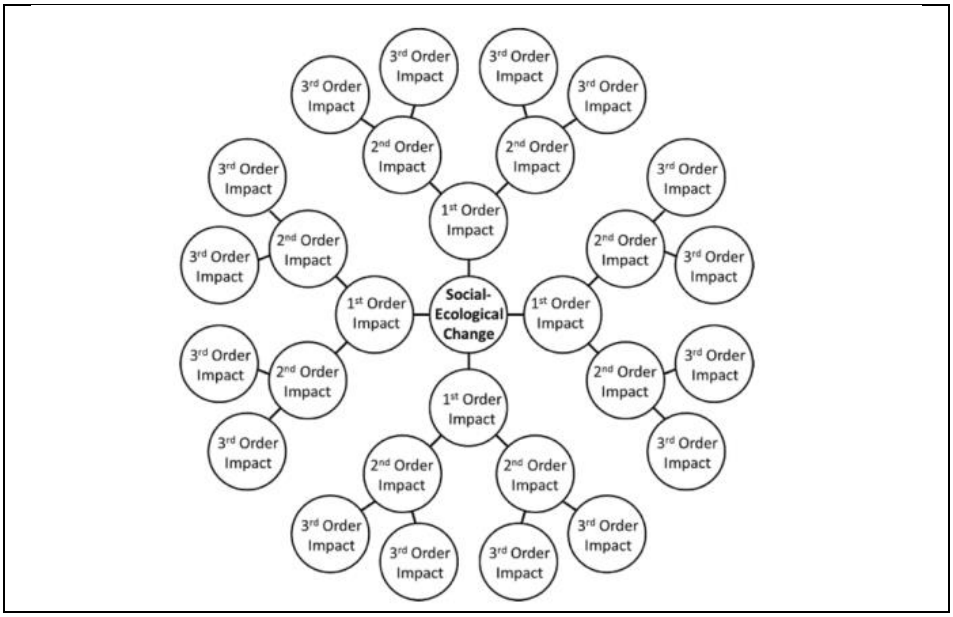

Figure 8: Simplified Futures Wheel structure (Bengston, 2016).

In the case of the Futures Wheel "*The process begins by carefully defining and describing the change to be examined… This could be a social trend affecting natural resources, an innovation in wildfire management, a new policy, or any significant change. Enough detail must be given for nonexpert participants to clearly grasp the nature of the change... Defining the center is done prior to the group process by the research or leadership team*" (Bengston, 2016)

However, there are significant differences in the way the tree structure develops. With the Futures Wheel there is no limit to the number of first-order branches added to the centre, or second-order branches added to first-order branches, et cetera. "*The number of first-order consequences generated is highly variable, but typically ranges from around 5 for a narrowly defined and relatively simple change, to 15 to 20 or more for a broad and complex change*" (Bengston, 2016). The net result is that the size of the tree structure will grow exponentially from one iteration to the next i.e. from centre outwards. The requirements of participants' time will grow accordingly. One example cited shows the numbers of consequences growing from 5 to 26 to 122. Not surprisingly the number levels of impact that are explored, is typically no more than three (Bengston, 2016).

In contrast, a ParEvo branching structure has a fixed number of surviving storylines, which is equal to the number of participants. The participants' time requirements remain the same from one iteration to the next. The differences between these two methods are akin to what is described in machine learning terms as the difference between depth-first and breadth-first types of search algorithms used for searching tree or graph data structures[9]. The Futures Wheel Is a breadth-oriented approach and ParEvo is a depth-oriented approach.

Another related difference, probably related to the time requirements, is that the content of each of the consequences statements appears to be limited to one sentence, rather than the paragraph as is the case with ParEvo. ParEvo storylines provide more of what anthropologist call 'thick description ' (Geertz, 1973). This distinction has implications for the types of knowledge that subsequently can be extracted, either through manual coding of content, or machine learning algorithms (e.g. topic modelling). In a recent ParEvo beta test exercise manual coding process led to the identification of 34 "problem" codes and twenty-eight "solution" codes, with half of these accounting for 82% of all coded instances. In contrast, in the Oteros-Rozas (2015) review of twenty-one participatory scenario planning processes, which took an inductive approach starting with the identification of drivers, "*…most commonly, ten or fewer drivers were identified*".

Like machine learning algorithms, both the Futures Wheel and ParEvo can include the use of an evaluation function to assess search results (i.e. the branch contents). Both appear to use similar evaluation criteria. The Implications Wheel, which is a refined version of the Futures Wheel developed by futurist Joel Barker, includes a group scoring process for desirability and likelihood (Schreier 2005). As with ParEvo, it is recognised that additional criteria may be relevant in different circumstances. Likewise, with ParEvo, a significant volume of material can be generated by a single exercise. "*The large number of possible consequences—often in the hundreds—combined with the ratings of importance, desirability, and likelihood can result in a significant amount of data. Inductive thematic analysis may be performed to identify broad themes*" (Benckendorff, 2008).

After the difference in search strategies, a second important difference between the two approaches is in the management of the participants. With ParEvo the way in which people participate is highly structured, and transparent. Anonymity of contributions is available to participants within the exercise, but identity of contributions is available to facilitators for subsequent analysis. With the Futures Wheel this situation appears to be the reverse. The generation of new consequences seems to be in the hands of a small group of participants (between 4 to 6), who are interacting with each other in the process of identifying new consequences. But there is no systematic process of documentation that might identify the individual sources of these contributions and make them available for subsequent analysis by other parties.

The third important difference between the two approaches is the role of theory. The core features of the ParEvo process have their origins in evolutionary epistemology (Davies, 1998). In addition, many of the ideas being proposed as being ways of analysing the nature of people's participation and its effects on the contents of exercises, are drawn from the more recent field of knowledge known as 'collective intelligence '. In contrast, the design of the Futures Wheel approach has relatively little basis in theory.

Challenges

Expertise

In many traditional approaches to scenario development and planning, expert knowledge is a key part of the scenario development process. In more participatory approaches there may need to be trade-offs. The Oteros-Rozas et al (2015) paper on participatory scenario planning examines a set of 23 case studies: "*Nine cases cited the lack of quantitative information, statistical and data-based testing, or modelling to support trends analysis as weaknesses. Five cases reported as a relevant weakness the unavoidable trade-off between the accuracy requested by the science base, which includes high complexity of scientific information, versus the social relevance of the process*"

In a ParEvo exercise there are different opportunities where expert knowledge can be brought into play:

- Participants can be given a pre-exercise briefing by relevant experts.
- Experts can be represented as one of the groups of stakeholders participating in an exercise,
- Experts can provide additional context-setting information, shared with all participants, and channelled through the Facilitator's updated guidance given at the start of each new iteration
- Experts can be involved as independent evaluators of surviving storylines, especially in relation to plausibility and probability of the described sequence of events.
- Experts can be involved in subsequent content analysis of all the storylines, surviving and extinct

These options have yet to be tested, but there are no technical obstacles preventing this from happening in future exercises.

A word of caution

The knowledge manifested in the results of a ParEvo exercise may not necessarily be very impressive, to the facilitator, other stakeholders or even the participants themselves. People have their limitations. It may often be the case that people do not know very much about a subject of concern. Perhaps it should not necessarily be expected that a ParEvo time process will always deliver impressive results, in the form of creative and collectively constructed scenarios?

If so, it may be useful to treat a ParEvo exercise also as a means of *explicating the limits* of a group's collective knowledge. An important aspect of metacognition is knowing what is not known (Angner, 2020). If so, this suggests that almost as a default procedure, a third party should always be brought in to examine what has been produced, and particularly to identify what is missing.

Scalability

Only eight ParEvo exercises had been carried out to date. The largest number of participants in any one exercise so far has been eleven. The larger the number of participants the wider the tree structure will be in the user interface, which may present technical problems. The other potential problem, which has been of greater concern, is the possibility that once the number of participants gets large no one will be able to read all the current storylines and thus make an informed choice.

But this may be a mistaken concern. As the number of participants and storylines grow, participants will have to resort to sampling storylines (if they have not done so already). In the absence of cues about authorship, this may well be a quasi-random process. If so, this may not be a bad thing.

Lorenz et al. (2011) have proved that the diversity of views within a group will decline when the group is fully exchanging information, and this was associated with poorer group performance on simple estimation tasks. So, if different individuals are taking different samples of the available information this trap might be avoided. There seems to be a parallel here with the use of "bagging" in ensemble methods of machine learning, where multiple random samples of observations in a data set are used to generate an aggregate prediction that is better than any model based on a single sample[10].

Other research suggests that the need to scale up, in order to achieve a satisfactory level of diversity in a group of participants, will not require large numbers. Vercammen and Burgman (2019) found that "*Based on overlap between confidence intervals, we can be reasonably confident that the collective IQ exceeds the average IQ within the group only if the group exceeds 6 members*". As the group size was increased, one at a time, the performance of the group increased but at a diminishing rate. By the time the group size reached 12 the performance of the group had reached 86% of its maximum possible (i.e. when the group size with increased up to 20).

Commissioners for a ParEvo exercise may have other reasons for expanding the number of participants in a ParEvo exercise. An organisation may have many different specialised units, but nevertheless want to see a set of scenarios that incorporates views from all those units. There are ways of doing so while not necessarily dramatically expanding the number of participants in a ParEvo exercise. The simplest approach is for each unit to have a single representative participating in the exercise. There is some evidence that this modular approach has other benefits. According to Pescetelli et al, (2020) "… *information aggregation across small homogeneous groups can be more effective in such environments, because it maintains efficiency within groups while preserving information diversity across groups*".

Another approach, currently under consideration for an exercise now being planned, is the equivalent of having multi-member constituencies in a parliamentary democracy. Anybody in an electorate can approach any one of the N elected members representing that electorate with their views and concerns. The second approach is being preferred because it is thought that it may help overcome, rather than reinforce, a "silo" mentality. Because of the way the ParEvo process works it will be possible to do an ex-post analysis of the extent to which silos were crossed, and how this affected the construction of storylines (which will vary in their probability, desirability, and other attributes).

In other circumstances organisers of a ParEvo exercise may want to canvas the opinions of a much wider group, who cannot be as easily organised. The other words, some form of crowdsourcing. Two possibilities have been identified but not yet tested. Both involve the use of queues. First option works very simply, based on "first come – first served", until the maximum number of desired participants per iteration has been reached. Second option is a variant, where participants in a previous iteration continue to participate unless their last contribution has not subsequently been added to (i.e. that storyline has become extinct). If not, they are replaced by the next person in the queue. In a sense this is a more radically Darwinian version of the current ParEvo design. Not only do certain ideas become extinct, but so do their carriers (metaphorically speaking). This approach has a connection to the often-quoted claim that "science progresses one funeral at a time "[11]. This claim is more than just a figure of speech, it is supported by recent research evidence (Azoulay et al, 2019).

References

Allington, G., Fernandez-Gimenez, M., Chen, J., & Brown, D. (2018). Combining participatory scenario planning and systems modeling to identify drivers of future sustainability on the Mongolian Plateau. Ecology and Society, 23(2). https://doi.org/10.5751/ES-10034-230209

Agner, E. (2020, April 13). Epistemic Humility—Knowing Your Limits in a Pandemic—By Erik Angner. Behavioral Scientist. https://behavioralscientist.org/epistemic-humility-coronavirus-knowing-your-limits-in-a-pandemic/

Barker, J. A. 2011. The Implications Wheel. http://implicationswheel.com/ (Retrieved 14 May 2020)

Bengston, D. N. (2016). The Futures Wheel: A Method for Exploring the Implications of Social–Ecological Change. Society & Natural Resources, 29(3), 374–379. https://doi.org/10.1080/08941920.2015.1054980

Bengston, D. N. (2019). Futures Research Methods and Applications in Natural Resources. Society & Natural Resources, 32(10), 1099–1113. https://doi.org/10.1080/08941920.2018.1547852

Benckendorff, P. 2008. Envisioning sustainable tourism futures: An evaluation of the futures wheel method. Tourism and Hospitality Research 8(1):25–36. doi:10.1057=thr.2008.2

Borgatti, S.P., Everett, M.G., & Freeman, L.C. (2002). Ucinet for Windows: Software for Social Network Analysis. Analytic Technologies. https://sites.google.com/site/ucinetsoftware/home

Buffardi, A. (2016). When theory meets reality: Assumptions, feasibility and implications of a complexity-informed approach. ODI. https://www.odi.org/publications/10432-when-theory-meets-reality-assumptions-feasibility-and-implications-complexity-informed-approach

Centre for the Evaluation of Complexity Across the Nexus (CECAN) (2020). Retrieved 11 May 2020, from https://www.cecan.ac.uk/

CECAN. (2016). Agent-Based Modelling for Evaluation A CECAN Evaluation and Policy Practice Note for policy analysts and evaluators. CECAN. https://www.cecan.ac.uk/sites/default/files/2018-01/HELEN%20ABM%20PPN%20v0.4.pdf

Coxon, A. P. M. (1999). Sorting data: Collection and analysis. SAGE.

Craven, L. K. (2017). System Effects: A Hybrid Methodology for Exploring the Determinants of Food In/Security. Annals of the American Association of Geographers, 107(5), 1011–1027. https://doi.org/10.1080/24694452.2017.1309965

Davies, R. (1998). Order and Diversity: Representing and Assisting Organisational Learning in Non-Government Aid Organisations [PhD, University of Wales - Swansea]. http://mande.co.uk/blog/wp-content/uploads/2013/05/thesis.htm

Davies, R., & Dart, J. (2005). The 'Most Significant Change' (MSC) Technique: A Guide to Its Use. http://www.mande.co.uk/docs/MSCGuide.pdf

Dennett, D. C. (1996). Darwin's Dangerous Idea: Evolution and the Meanings of Life. Penguin UK.

Derbyshire, J. (2019). Use of scenario planning as a theory-driven evaluation tool. FUTURES & FORESIGHT SCIENCE, 1(1), e1. https://doi.org/10.1002/ffo2.1

Dobzhansky, T. (1973). Nothing in Biology Makes Sense except in the Light of Evolution. The American Biology Teacher, 35(3), 125–129. https://doi.org/10.2307/4444260

Frith, D., & Tapinos, E. (2020). Opening the 'black box' of scenario planning through realist synthesis. Technological Forecasting and Social Change, 151, 119801. https://doi.org/10.1016/j.techfore.2019.119801

Kobes, S., & Loy, T. R. (2020). Whatever Happened to Scenario Planning? A Systematic Literature Review (SSRN Scholarly Paper ID 3536419). Social Science Research Network. https://doi.org/10.2139/ssrn.3536419

Geertz, C. (1973). The interpretation of cultures: Selected essays. Basic Books.

Global Learning for Adaptive Management initiative (GLAM). (2020). ODI. https://www.odi.org/projects/2918-global-learning-adaptive-management-initiative-glam

Nilsson, A. E., Carson, M., Cost, D. S., Forbes, B. C., Haavisto, R., Karlsdottir, A., Larsen, J. N., Paasche, Ø., Sarkki, S., Larsen, S. V., & Pelyasov, A. (2019). Towards improved participatory scenario methodologies in the Arctic. Polar Geography, 0(0), 1–15. oh kitchensinkhttps://doi.org/10.1080/1088937X.2019.1648583

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. Proceedings of the National Academy of Sciences, 108(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108

McBride, M., Lambert, K., Huff, E., Theoharides, K., Field, P., & Thompson, J. (2017). Increasing the effectiveness of participatory scenario development through codesign. Ecology and Society, 22(3). https://doi.org/10.5751/ES-09386-220316

Mitchell, S. D. (2009). Unsimple Truths: Science, Complexity, And Policy. University of Chicago Press.

Oteros-Rozas, E., Martín-López, B., Daw, T., Bohensky, E., Butler, J., Hill, R., Martin-Ortega, J., Quinlan, A., Ravera, F., Ruiz-Mallén, I., Thyresson, M., Mistry, J., Palomo, I., Peterson, G., Plieninger, T., Waylen, K., Beach, D., Bohnet, I., Hamann, M., … Vilardy, S. (2015). Participatory scenario planning in place-based social-ecological research: Insights and experiences from 23 case studies. Ecology and Society, 20(4). https://doi.org/10.5751/ES-07985-200432

Page, S. E., Lewis, E., Cantor, N., & Phillips, K. (2017). The Diversity Bonus: How Great Teams Pay off in the Knowledge Economy. Princeton University Press.

Planque, B., Mullon, C., Arneberg, P., Eide, A., Fromentin, J.-M., Heymans, J. J., Hoel, A. H., Niiranen, S., Ottersen, G., Sandø, A. B., Sommerkorn, M., Thébaud, O., & Thorvik, T. (n.d.). A participatory scenario method to explore the future of marine social-ecological systems. Fish and Fisheries, 0(0). https://doi.org/10.1111/faf.12356

Ramirez, R., & Brodhead, D. (2019, September 3). How can scenario planning help evaluate the impact of research? Evaluation and Communication in Practice. Retrieved from https://evaluationandcommunicationinpractice.net/how-can-scenario-planning-help-evaluate-the-impact-of-research/

Reinhardt, J., Liersch, S., Abdeladhim, M., Diallo, M., Dickens, C., Fournet, S., Hattermann, F., Kabaseke, C., Muhumuza, M., Mul, M., Pilz, T., Otto, I., & Walz, A. (2018). Systematic evaluation of scenario assessments supporting sustainable integrated natural resources management: Evidence from four case studies in Africa. Ecology and Society, 23(1). https://doi.org/10.5751/ES-09728-230105

Schreier, J. (2014). Evaluating a Simulation with a Strategic Exploration Tool. Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL Conference, 32(0), Article 0. https://journals.tdl.org/absel/index.php/absel/article/view/629

Seeve, T. (2018). A Structured Method for Identifying and Visualizing Scenarios [Aalto University]. https://pdfs.semanticscholar.org/d7cf/81104a41ca4e1b0e590ceba48823785c07e1.pdf?_ga=2.224229203.915126767.1584870457-791828924.1584870457

Spradley, J. P. (1979). The Ethnographic Interview. Fort Worth, Tex. : Harcourt Brace Jovanovich College Publishers.

Totin, E., Butler, J. R., Sidibé, A., Partey, S., Thornton, P. K., & Tabo, R. (2018). Can scenario planning catalyse transformational change? Evaluating a climate change policy case study in Mali. Futures, 96, 44–56. https://doi.org/10.1016/j.futures.2017.11.005

Vercammen, A., Ji, Y., & Burgman, M. (2019). The collective intelligence of random small crowds: A partial replication of Kosinski et al. (2012). Judgment and Decision Making, 14(1), 91–98.

Vogel, I. (2012). Review of the use of 'Theory of Change' in international development. DFID. http://www.dfid.gov.uk/r4d/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf

Young J, Susan Njambi-Szlapka, & Janise Rodgers. (2019). Practical science for uncertain futures: Using scenarios to improve resilience to earthquakes. ODI. https://www.odi.org/publications/11469-practical-science-uncertain-futures-using-scenarios-improve-resilience-earthquakes

Yu, C., Chai, Y., & Liu, Y. (2018). Literature review on collective intelligence: A crowd science perspective. International Journal of Crowd Science, 2(1), 64–73. https://doi.org/10.1108/IJCS-08-2017-0013

[1] The Oteros-Rozas (2015) review of 21 participatory scenario planning exercises is a good example, which differentiates cases by ecological zone, governance context, economic zone and livelihood. On a larger scale. there are clearly identifiable types of scenario planning that can be described as specialist, and only use by particular organisations or types of organisations, and others which are clearly more generalist, being used by a wide variety of organisations. An example of the first is the widely referenced scenario planning method first used by Shell (reference). An example of the second type would be Delphi (reference)

[2] http://www.aptivate.org/

[3] https://mscinnovations.wordpress.com/

[4] https://mscinnovations.wordpress.com/introduction/ten-stages/

[5] Content analysis methods and results at: https://mscinnovations.wordpress.com/evaluation/ and participation analysis methods and results at: https://mscinnovations.wordpress.com/measures/

[6] See https://mscinnovations.wordpress.com/introduction/ten-stages/#using

[7] See https://en.wikipedia.org/wiki/Process_tracing

[8] See https://mande.co.uk/special-issues/evolving-storylines-a-participatory-design-process/

[9] See https://en.wikipedia.org/wiki/Breadth-first_search and https://en.wikipedia.org/wiki/Depth-first_search

[10] See https://en.wikipedia.org/wiki/Bootstrap_aggregating for an explanation of bagging

[11] See https://en.wikipedia.org/wiki/Planck%27s_principle