*Article*

# ParEvo: A methodology for the exploration and evaluation of alternative futures

**Rick Davies** iD
Independent Evaluation Consultant, Cambridge, UK

**Tom Hobson** iD
Centre for the Study of Existential Risk, Cambridge, UK

**Lara Mani** iD
Centre for the Study of Existential Risk, Cambridge, UK

**Simon Beard** iD
Centre for the Study of Existential Risk, Cambridge, UK

## Abstract

Evaluators' main encounter with views of the future is in the form of theories of change, about how a programme will work to achieve a desired end, in a given context. These are typically focussed on specific relatively short-term futures, which are both desired and expected. But even in the short term, reality often involves unpredictable events which must be responded to. Other ways of thinking about the future may be helpful and complementary, notably those developed by foresight practitioners working in the field of futures studies. These pay more attention to a range of possible futures, rather than a single perspective. One way of exploring such futures is by using ParEvo.org, an online process that enables the participatory exploration of alternative futures. This article explains how the ParEvo process works, the theory informing its design, and its usage to date. Attention is given to three evaluation challenges, and methods to address them: (a) optimising exercise design, (b) analysis of immediate results and (c) identifying longer-term impacts. Two exercises undertaken by the Cambridge-based Centre for the Study of Existential Risk (CSER) in 2021–2022 are used as illustrative examples.

## Keywords

evaluation, futures, online, ParEvo, participatory

**Corresponding author:**
Rick Davies, 82 Bishops Road, Cambridge CB2 9NH, UK.
Email: rick.davies@gmail.com

'It is clear from both evolutionary and immunological theory that in facing an unknown future, the fundamental requirement for successful adaptation is pre-existing diversity'

—Edelman (1987)

## How do evaluators think about the future?

Evaluators think about the near future in different ways. They are familiar with the idea, and value, of having an explicit theory of change about the future. When completing an evaluation, they will typically be expected to make recommendations about what should be done in the near future (Feinstein, 2019). Less frequently, they may be involved in assessments of the evaluability of an intervention (Austrian Development Agency (ADA), 2022).

One common characteristic of these three different ways of engaging with the future is their *short-term* nature. Events beyond the expected lifespan of an intervention do not get much attention. According to a quick search of the UK government website Development Tracker, the average lifespan of their development aid interventions is around 4 years.[1]

There are other limitations. Theories of change, as encountered by evaluators, or reconstructed by them, embody a particular way of seeing the future. To be useful, they need to describe futures which are both *desired* and *expected* to be achievable. They are optimistic predictions. Undesirable and unexpected futures that might derail these expectations are given more limited attention either in the margins of diagrams or in separate tables, dealing with risk and assumptions.

They also tend to represent relatively *singular* views. While more sophisticated diagrammatic representations of theories of change may include multiple intersecting pathways connecting activities to outcomes, they are in most cases *convergent* models, oriented to a specific view of the future (Davies, 2018).

### Coping with unpredictability

'No battle plan survives contact with the enemy', it was once said.[2] Theories of change also have their recognised limitations. Almost all widely cited guidance texts on the use of theories of change repeat the advice that theories of change need to be updated in the light of experience (Funnell and Rogers, 2011; Rogers, 2014). However, doing so is easier said than done, because of the time and resources that can be involved in articulating and agreeing on a theory of change, and its revisions.

The unpredictability of reality is also a common theme of writers on complexity and its implications for evaluations and evaluators (Morell, 2022). One response is the attempt to develop more sophisticated models of interventions and their contexts, representing complex networks of actors, events, and their relationships. Both in the form of static systems models and dynamic simulations (Barbrook-Johnson and Penn, 2022).

Parallel to both developments, for some decades now, have been efforts to develop more adaptive forms of management, which can respond to and exploit unexpected events (Rogers and Macfarlan, 2020; Starr and Peek, 2019). If these efforts do include a perspective on the future, it is one where the future consists of multiple small experiments, in series and in parallel, each containing a conjecture about what might work in specific circumstances. It is more short-term, but less singular, in vision.

Evaluators thinking about theories of change, complexity and adaptation have all been influenced by other disciplines. But perhaps not as much as might be expected. Apparently, beyond the horizon of many evaluators, decades of experience have already accumulated with a range of methods for exploring and thinking about the future, as seen in papers published by journals such as *Futures: The Journal of Policy, Planning and Futures Studies*, first published in 1968. More than 20 peer-reviewed journals now cover the field (World Futures Studies Federation (WFSF), n.d.). They not only cover conjectures about different futures, but also a wide variety of methods for exploring possible and likely futures. The latter is what is of interest here, because of their potential relevance to evaluators, and the designers of interventions they evaluate.

Many of these methods have been already documented in manuals produced by agencies that can also be commissioning evaluations (Organisation for Economic Co-operation and Development (OECD), n.d.; UNDP, 2018; Waverley Consultants, 2017) Some well-established methods, such as the Delphi technique, are logistically demanding and likely to have a limited range of appeal. Others, such as the Futures Wheel (Bengston, 2016), have more immediate potential for enabling programme developers to think more comprehensively about the many different possible consequences of an intervention. ParEvo is a new contribution to the menu of available methods which we think will appeal to both evaluators and foresight practitioners.[3]

## Enabling the participatory exploration of alternative futures

In 2019, the lead author developed a web application called ParEvo, designed to enable the participatory exploration of alternative futures, online. Not to predict the future, but to identify a range of futures that may need to be thought about. The design of the process of participation was based on what has been called the evolutionary algorithm (Dennett, 1996: 48). That is, the reiteration of variation, selection, and retention of forms of information. This choice had its origins in an earlier PhD research on organisational learning, which also led to the design of another now widely used participatory method known as Most Significant Change (MSC) (Davies, 1998; Davies and Dart, 2005), based on the same epistemology. Now used widely for impact monitoring and evaluation in complex development projects. The two methods have complementary characteristics. While MSC looks backwards and uses a convergent and optimising process to identify significant changes, ParEvo looks forwards and is more divergent and satisficing (Simon, 1956), as will be explained below.

### How a ParEvo exercise works

*The generation of storylines.* Via an online interface at ParEvo.org, participants are presented with a seed text equivalent to the first paragraph of a novel, which is prepared by the exercise facilitator. Participants are then each given the opportunity, independently and anonymously in parallel, to extend that narrative by adding a following paragraph, describing what happened next. Once all these contributions are received online, participants are then enabled to view each of those alternative extensions. In the next iteration, they then choose one of those contributions which they would most like to develop by adding another following paragraph, again independently and anonymously in parallel. The exercise then repeats this sequence over a series of iterations. In most exercises, one immediate result will be that some initial versions of the story will be ignored, while others will be extended by one or
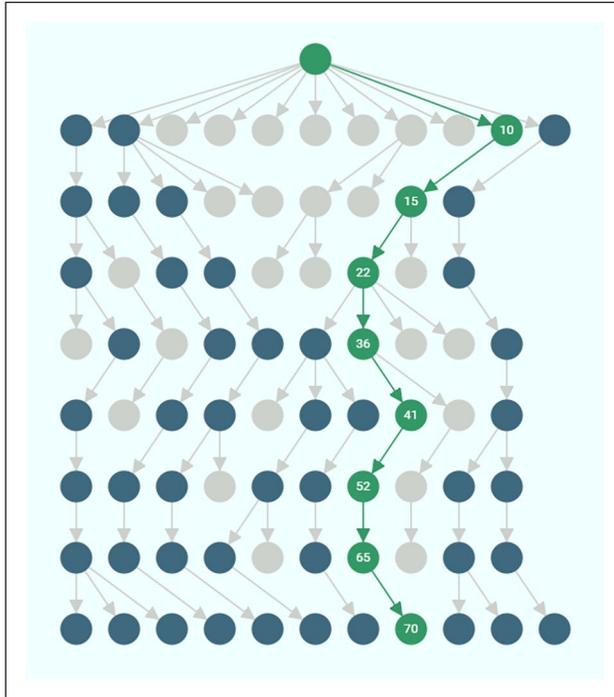
**Figure 1.** Tree structure of alternative storylines generated by CSER exercise 1: Biological Research.

more participants. As this process is reiterated what emerges is a tree structure of alternative storylines (Figure 1).

In this example, discussed in more detail below, there were up to 11 participants, who participated in eight iterations, visible as rows. The process begins at the top with a single seed paragraph and ends at the bottom, with 11 surviving storylines. Each node represents a paragraph of text contributed by a participant, and the connecting lines show which new paragraph was added to which pre-existing paragraph. Grey nodes in the tree structure indicate paragraphs which were not continued and which in effect represent extinct storylines. Dark green nodes represent paragraphs which others did build on and which became part of storylines which survived until the end of the exercise (bottom row). Bright green nodes represent one storyline which has been highlighted by a user of the ParEvo app. Doing so then brings up the full text of that storyline in a panel to the right of the tree structure, on the ParEvo user interface (not shown above).

*Evaluation.* When the process of generating a set of storylines is finished participants are then asked to evaluate the storylines that they have helped to generate. This can be done using the widget built into the app and/or with the use of a more detailed online survey. The widget allows participants to make polar evaluation choices, of which storyline they think is most or least likely, desirable, equitable, sustainable and so on. The available choices are set by the exercise facilitator. The online surveys used to date have used both open- and closed-ended questions about the contents and process. In addition, the facilitator can download from ParEvo.org 12 Excel formatted datasets containing automatically generated information about

the contents of the storylines and the structure of people's participation in the exercise. Exercise facilitators have also organised post-exercise meetings of participants to share and discuss responses to the evaluation survey, analyses of downloaded data and other perspectives on the exercise as seen by the participants.

Many of the methods of evaluating the products and process of a ParEvo exercise were developed after the initial development of the web interface enabling participants to generate a range of alternative storylines about the future. In the process, it has become increasingly evident that the post-exercise evaluation stage is equally important because it enables both the participants and the facilitator to extract more value from both the storylines that have been generated and the process used to generate them.

*Types of theories.* There are three different types of theories of change about what happens during a ParEvo exercise. The first is the participants' own often tacit and informal theories about what might happen in the future, as then evident in the contents of their contributions, and the resulting composite storylines.

The second is the exercise facilitator's expectations of what they want to see happen with the exercise they have designed and organised. As evident in the kind of futures they want to see explored, the kinds of people to be involved, the time span and granularity of the exercise and the guidance they give to participants at the beginning of each new iteration. The expectations and findings of the CSER staff who facilitated the two ParEvo exercises described later below will be the subject of other forthcoming publications by Mani, Hobson and Beard.

The third is the expectations of the platform administrator (and designer of the ParEvo app and lead author of this article). Each consecutive exercise has been viewed as an opportunistic experiment, usually involving some new variations in the design parameters, primarily under the control of the exercise facilitator – not the administrator. Some of these parameters are stable across multiple exercises and while others vary. In addition to the evolutionary epistemology at the base of the process design used by all, there is an associated ongoing interest in the role of diversity. This perspective has been informed by writings on diversity from a complexity perspective (Page, 2008; Page et al., 2017), measures of diversity used in ecology and sociological uses of those ecological ideas (Stirling, 2007) and network analysis as a way of visualising and measuring diversity measures (Borgatti et al., 2018).

Considerations about appropriate measurement methods have arisen when thinking about diversity as an independent variable affecting the creativity of the process but also as a dependent variable that is descriptive of the range of possible futures that have been explored. Underlying the design of ParEvo, when used to look forward, is the assumption that the generation and analysis of a diversity of storylines will enable participants to be better prepared for the future, which is only likely to be partially knowable at best. In this context, the intention is not to predict the future but to be able to be more adaptive and responsive to the futures that might take place. This approach is consistent with a substantial body of evidence on the importance of diversity to the more general task of effective problem solving, as referred to recently in Campbell et al. (2022).

## Past and future uses of ParEvo

As of mid-2022, 19 different ParEvo exercises have been completed, during and since the development of the app.[4] Participants have included school students, volunteers recruited

from M&E communities of practice, crowdsourced paid adult university-educated UK partici-pants, staff from a UK development aid think tank, UN Volunteers, UN agency staff members and internationally recognised experts in particular scientific fields. Futures explored include Post-Brexit Britain, Climate change post-COP26, post-Trump USA, a 5-year corporate strat-egy and possible uptake pathways for educational research. Alternative histories have also been explored, including agricultural development project implementation, UNV volunteer experiences and gender policy implementation within a UN agency. Eight of the earlier exer-cises were initiated by the lead author, and 11 of the more recent exercises were initiated by members of other interested organisations. Two of these were implemented by the Centre for the Study of Existential Risk (CSER), in Cambridge, UK, and are used as illustrative exam-ples in this article.

ParEvo is a substantially different way of thinking about the future than most evaluators will be familiar with. While there are some other online platforms they can use to think about theo-ries of change (Clark, 2022), these operate within the framework and associated limitations spelled out in the first section of this article. They do not focus on the collaborative develop-ment and evaluation of a *diversity* of views about possible futures. Other online platforms have been developed by futurists for the exploration of the future, which was reviewed by Raford (2015). ParEvo exercises differ from these in at least five respects. First, there is a generative theory informing the process design – the social embodiment of the evolutionary algorithm. Second, the process is more divergent and satisficing rather than convergent and optimising, as is the case with online Delphi exercises – perhaps the most widely known method. Third, with ParEvo, the construction of narratives precedes and informs a detailed analysis, rather than fol-lowing and being informed by a technical analysis of other available data. In this respect, it is more ethnographic in orientation, taking participants' views as the primary resource material. Fourth, the greater diversity of alternative futures compared to the average futures exercise, stereotypically represented in a 2 × 2 matrix (Rhydderch, 2017). Finally, the greater promi-nence that it gives to the evaluation component of the whole process.

As of mid-2022, four other organisations had expressed interest in using the ParEvo pro-cess. A Dutch NGO plans to use ParEvo as an alternative approach to strategic planning about the future direction of their organisation. A European NGO network wants to explore future possibilities for the development of civil society. A humanitarian network will be testing its grant-making quality assurance framework against a range of imagined futures. An international NGO intends to explore alternative futures involving strengthening social movements on gender equality and sexual and reproductive rights in Europe. In addition, one existing user will be repeating their ParEvo exercise on an annual basis, exploring how views on uptake pathways for educational research develop over time, in a specific community of researchers.

## The CSER exercises

In 2021 the CSER in Cambridge, UK, expressed interest in using ParEvo.org to explore alter-native futures relating to different forms of existential risk. These are. . . 'risks that could lead to human extinction or civilisational collapse' (CSER, 2022). These include climate change, nuclear winter, pandemics, super volcanoes, unconstrained artificial intelligence, and other possible harms caused by emerging technologies, including biological research (Cassidy and Mani, 2022; Christian, 2020; Hobson et al., 2022; Kemp et al., 2022; Maas et al., 2021; Witze,

2020). Although largely focussed on undesirable futures, there is also a significant thread through the associated literature that focuses on describing desirable futures and ways in which 'human flourishing' and other states can be achieved.

In mid-2021, a pre-test was carried out, using a small group of staff from within the Centre. In late 2021, the first full-scale exercise explored 'Future Pathways for Governing Biological Research', with a focus on the development of global mechanisms for the management of biological research risks.[5] The facilitators were Tom Hobson and Lara Mani, both Research Associates at the Centre. Tom Hobson, being a specialist in this risk area, recruited 11 internationally recognised experts in biological research risks. Lara Mani's immediate interest was in the comparative merits of ParEvo versus game-based approaches for exploring possible futures. Specifically, Intelligence Rising, a role-playing game that had already been developed by another CSER research associate (Avin et al., 2020).

In May 2022, a second exercise explored futures relating to 'Developing the science of global risk'.[6] This was facilitated by S J Beard, Academic Programme Manager and Senior Research Associate at the Centre. Participants in the exercise were 10 mainly early-career researchers in the field of existential risk, participating in an Intensive Visitor Programme (IVP) following an international conference hosted by the Centre. Participants in this exercise did meet face-to-face daily, as part of the IVP but interacted with the ParEvo app independently and anonymously online in their own time. In contrast, in the first exercise on biological research risks, the participants did not meet together until after the evaluation stage.

Diagrammatic representations of the storylines generated in these two CSER exercises are shown in Figure 2 below. Key quantitative features of the two exercises are summarised in Table 1 below.
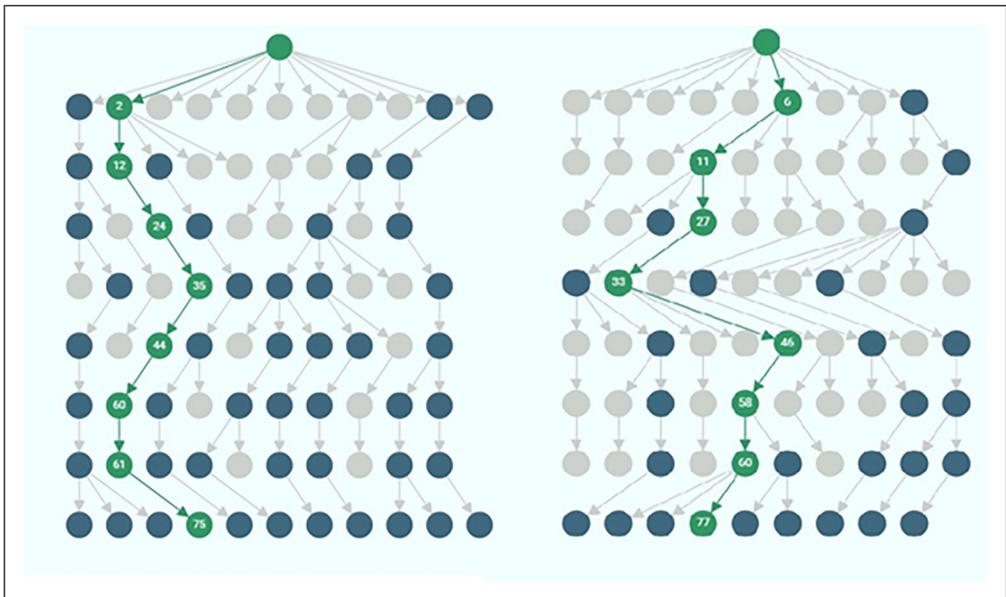


**Figure 2.** Tree structures for CSER exercises 1 (left) and 2 (right).

**Table 1.** Quantitative features of the two CSER exercises.

|  | Exercise 1 Biological research | Exercise 2 Global risk research |
|---|---|---|
| Max. # of participants | 11 | 10 |
| Min. # of participants | 6 | 9 |
| # of iterations | 8 | 8 |
| # of contributions | 80 | 77 |
| # of surviving storylines | 11 | 9 |
| # of extinct storylines | 20 | 25 |
| Approximate total word length | 12,000 | 11,550 |

## Evaluation challenges and responses

Evaluation opportunities arise at various points before, during and after ParEvo exercises. How those opportunities are used is largely in the hands of each exercise facilitator, bearing in mind their aims for their own exercise. Some of these expectations are informed by information shared by the ParEvo platform administrator, which has been accumulated from previous exercises.

From the platform administrator's point of view, there are three generic evaluation challenges that are relevant to most exercises, relating to:

- Optimising the design of a ParEvo exercise.
- Usefully characterising the results of a ParEvo exercise.
- Assessing the subsequent impact of a ParEvo exercise.

Each of these is discussed at length below.

### *Optimising design*

When designing an exercise, a facilitator needs to make choices about multiple parameters, describing both independent and dependent variables.

Independent variables describe the exercise settings. These include: the number of participants, the types of participants, the roles they have within an exercise (participant, commentator, evaluator, observer), the contents of the seed text, the number of iterations, the duration of time covered by each iteration, the maximum length of text for any contribution, the content of the guidance text given at the start of each iteration and the facilitator's use of the comment facility during the exercise. In the exercises to date, the greatest variation has been present in the types of participants and in the contents of the seed text and subsequent guidance given by the facilitator, that is, the qualitative variables. Within each exercise, facilitators typically update the contents of the guidance text with each new iteration. They can also use the comment function to influence participants' thinking – although to date this option has been used with caution. Within the context of all these settings, participants then make further choices, about whose contributions build on and the contents of their contributions, with each new iteration.

Dependent variables describe expected outcomes. These are of two types: those specific to each exercise – as defined by the facilitator and those which are more generic and likely to be applicable to many exercises – as identified by the administrator. The latter are of concern

here, the former are discussed briefly further below. As mentioned in the section on theory above, the generation of a diversity of alternative futures has been a default expected outcome for almost all ParEvo exercises to date. Drawing from the field of ecology, Stirling (2007) differentiated three facets of diversity, each of which are measurable:

- *Variety*, also known as richness, which is the number of different kinds, for example, species.
- *Balance*, also known as evenness, the relative numbers of each kind.
- *Disparity*, the degree of difference between kinds, for example, between people and chimpanzees versus people and bacteria.

In the analysis of ParEvo exercises, these aspects of diversity have been measured in three ways.

*Tree structures.* The first method looks at the network structure of the storylines, in terms of disparity. Some storylines are more similar than others, in that they share a large amount of common content, only diverging in the most recent iterations. The content of others is less so because they diverged in the earlier iterations. Comparing the structure of storylines in CSER exercises 1 and 2, we can see that whereas in Figure 2 exercise 1 four of the original storylines in iteration 1 survived to the last iteration, but only two did so in exercise 2. This aspect of diversity can be measured more specifically by counting the links connecting the surviving storylines, a simple network analysis measure of distance. There were 53 in exercise 1, versus only 30 in exercise 2. In exercise 2, after iteration 4, there was such a loss of diversity that the facilitator was concerned that only one original storyline might be left within the next iteration of two. And that a more directive form of guidance might need to be given in the next iteration to prevent this. In practice, this was not done, and the level of diversity, in terms of disparity, did recover thereafter without intervention. The significance of disparity is discussed further below, in the discussion of exploration versus exploitation strategies.

*Combinations of sets of ideas.* The second approach looks at the kinds of combinations of ideas that occurred, as a percentage of all possible combinations. Each participant can be seen as a set of ideas. When one participant's contribution builds on the contribution of another, this set intersection represents one of those kinds of combinations taking place. The total number of possible types of combinations in any exercise equals the number of participants times the number of participants.[7] The percentage of those possible combinations that did occur in exercise 1 was 61%, whereas in exercise 2 it was 70%. This measure, also known as network density, is a crude measure of 'variety' diversity.[8] This measure is of interest because the recombination of ideas is considered an important source of creativity both in biological evolution and human culture (Xiao et al., 2022).

*Storyline evaluations.* A third approach looked at diversity in the evaluation judgements of participants. At the end of both CSER exercises, after eight iterations had been completed, participants were asked to identify which specific surviving storylines they saw as describing the 'most likely', 'least likely', 'most desirable', and 'least desirable' futures – as seen from their own perspective. Their responses were then used to create a scatter plot within which there were four different quadrants of possibilities, as shown in Figure 3 below.
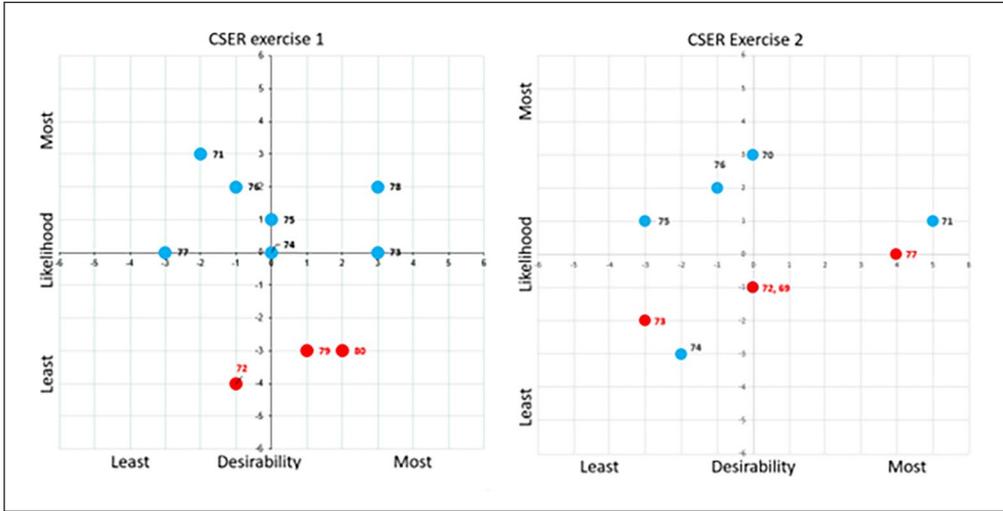
**Figure 3.** Distribution of evaluations of the surviving storylines, in CSER exercises 1 (left) and 2 (right). Numbered data points refer to individual storyline IDs.

The numbered nodes represent the surviving storylines that were identified as having one more attributes (i.e. more/less – desirable/likely). The values on each axis represent the number of participants who made judgements of that kind.[9] The assessment of storylines located towards the edges of the scatterplot had the support of more participants than those towards the centre. The red storylines were those where participants had very conflicting judgements, where both polar extremes were applied to the same storyline, for example, being most and least desirable. In exercise 1, two of the three contradictory judgements were about desirability. In exercise 2, three of the four contradictory judgements were about likelihood.

Diversity in this context can be measured in two different ways. Minimal (variety) diversity of judgement would be visible in the presence of only two storylines in the scatter plot, when all participants agreed that one storyline was most desirable and least likely, and the other was least desirable and most likely.[10] Maximum diversity would be visible where all the surviving storylines appeared on the scatterplot. In both of the CSER exercises, there was maximum diversity on this variety measure of diversity. Within all the possible combinations of judgements, the most disparate would be where participants expressed contradictory judgements about the desirability, or the likelihood, of a storyline. As noted above, these kinds of judgements were seen in both exercises, slightly more so with exercise 2.

*Implications for optimisation.* The diversity measurement options just discussed can be seen as mediating variables possibly affecting the post-exercise impacts exercise facilitators are aiming for. Or as dependent variables, of interest as more proximate outcomes. In both cases, the exercise settings can be seen as the independent variables. The relationship between exercise settings and the diversity of outcomes and their subsequent impacts is yet to be systemically explored. Ideally, findings will inform how future facilitators can optimise the design of their exercises.
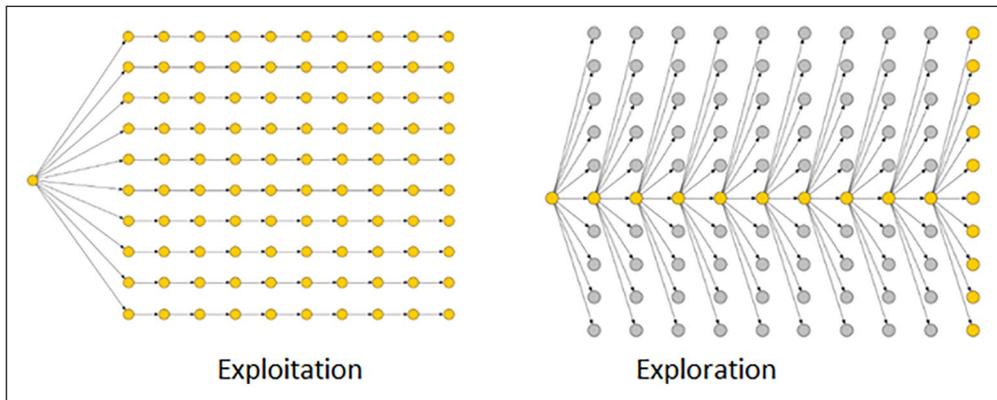
**Figure 4.** Extreme examples of exploitation versus exploration search strategies.

One dimension of optimisation is captured by the distinction made by organisational learning theorist James March (1991), between 'exploration' – of the new, and 'exploitation' – of what is already known. The question of which strategy is most appropriate in what conditions has been the subject of ongoing research since then (Wilden et al., 2018). The same distinction can be used to differentiate search processes used in different ParEvo exercises. The two extreme manifestations of these are shown in Figure 4. Exploration, in this setting, involves the development and then abandonment of many different storylines, whereas exploitation involves the more persistent development of a single preferred storyline.

Viewed from this perspective, exercise 2 participants seemed to be pursuing a slightly more exploratory futures search than those in exercise 2. Twenty five storylines were abandoned in exercise 2 versus 20 in exercise 1. Looking more broadly, all completed exercises to date have been on various points on the spectrum from exploration to exploitation.

In the future, it is possible that some facilitators of ParEvo exercises may want to give more emphasis to exploitation, and for their participants to converge on a more specific view of the future – in ways that are not necessarily evident in a particular tree structure. Two ways of enabling this have been identified. One is for the facilitator to give explicit guidance to participants that they should extend storylines which fit $\times$ criteria. At present, the default guidance is to allow participants to choose according to criteria of interest to themselves, albeit within the overall ambitions of the exercise.

The second option is to make use of a widget recently built into ParEvo which can provide participants with anonymised feedback on the extent to which their contributions have been built on by others (Figure 5). Receipt of such comparative feedback could motivate more conceptualy convergent contributions. That possibility is yet to be tested.

### Characterising the contents of storylines

The storylines generated by both CSER exercises produced a large volume of text, between 11 and 12,000 words each, constructed using 77 to 80 individual contributions. Characterisation involves identifying features of these storylines that may have some significance. Either cognitively – for how participants and other exercise stakeholders think about the future, or behaviourally – how they can respond in the present, near or distant future.

**Figure 5.** Leader board providing participants with feedback on responses to their contributions.

Characterisation has been done at three different scales of analysis: individual contributions, whole storylines, and the whole set of storylines. In the two CSER exercises, this was done by the participants themselves – via the evaluation survey, and then by observers and the facilitators – using survey and downloaded data. The process has been challenging. While a string of numbers can be summarised using a few measures such as mean, range and standard deviation, a long string of text, for example, a storyline is much more multidimensional. Even more so is a set of branching storylines. As described below, a wide range of approaches are being explored.

### Differences between contributions

*Content analysis coding.* Differences between contributions were identified through different forms of content analysis carried out by the facilitators, and by the administrator. In both exercises, facilitators did manual coding of themes in contributions by the facilitators, based on their domain knowledge. The number of codes used ranged from 18 in exercise 1 to 35 in exercise 2. While each schema does need to fit the specific exercise purpose, there is also some potential value in promoting the use of some common categories of themes, such as contexts, actors, activities and outcomes that might have a generic usefulness, for example, in the development and/or updating of theories of change.[11]

Automated tools have been useful. In exercise 1, text extraction software was found to be a useful way of automatically extracting words used to describe locations, persons or organisations (Text2Data.com, 2022). The locations of more than 80 different organisations were identified across the 80 contributions, far more than would feature in most theories of change.

Simpler semi-automated approaches were also found to be useful. In exercise 1, and other more recent exercises, the search facility built into ParEvo was used to complement automated text extraction with searches for country names. A dataset was then downloaded showing in matrix form the occurrence of searched words across all the contributions. Thirty different countries were mentioned in the exercise 1 storylines.

Semi-participatory approaches are also an option, not yet used. The ParEvo user interface includes the option for participants, and others, to apply 'tags' to individual contributions at the end of each iteration. 'Tags' are single-word labels, available from a menu pre-defined by the facilitator, such as 'desirable', 'unlikely', 'surprising' and so on.

*Analysis.* In both exercises, matrices containing contributions as cases and themes as attributes of those cases have been analysed in simple terms of frequencies and averages. For example, in exercise 1, comparisons were made of the relative frequency of mentions of different drivers of change.

*One-to-one relationships* were explored through cross-tabulations of themes against surviving versus extinct status of storylines, and the presence of specific storylines, to identify useful descriptors (See Beard et al., 2023).

Network visualisation tools (Borgatti et al., 2002) were also used to make more complex *many-to-many* interrelationships between themes more visible. In exercise 1, differences could be seen between named countries densely connected in the core versus more scattered and weakly connected groups on the periphery.

Simple predictive modelling software (EvalC3, 2022) was used to explore *many-to-one* relationships. In exercise 2, the whole set of themes was explored to identify if any combinations were good predictors of the presence of whether storylines survived, or not. No themes, or groups of themes, were found which were necessary or sufficient, but a 'probabilistic' model with a combination of 4 coded themes was found. The most predictive theme among these was the presence in the storylines of 'Individual researchers as agents of change'.[12]

Other forms of many-to-many analysis were combined with many-to-one analyses. In exercise 2, the cluster analysis method known as Topic Modelling was used to find what sets of words used within the contributions were good predictors of surviving versus extinct storylines (Boyd-Graber et al., 2017; The BigML Team, 2022). One merit of this approach was that it used the whole corpus of text, rather than those only found by automated text extraction and/or manual coding. Three clusters, known as topics, were found, each with a set of 10 words that were more likely to be found together than in the other clusters (Table 2).

Topic ratings assigned to each contribution were cross-tabulated against the survival status of the storylines. None of the topics predicted the survival of storylines with more than 56% classification accuracy – little better than chance. Nor did any combinations of the topics. While this method may be more useful with storylines generated by other exercises, or with different types of outcomes in mind, one problem seems likely to remain. This is how to meaningfully and justifiably name the clusters that have been identified. Even so, the individual words making up each cluster may still provide a useful source of additional theme codes in themselves. Such as the actor types underlined in Table 2.

*Differences between storylines.* Differences between storylines were identified using two different methods. The first involved criteria of difference pre-identified by the facilitators. The second involved more emergent criteria identified by each participant on their own.

**Table 2.** Topic models found by BigML, from exercise 2 contributions.

| Cambridge | Research | Risk |
| --- | --- | --- |
| Cambridge | Research | Risk |
| Policy | Community | Public |
| Existential | Increasingly | Ai |
| Scholars | Catastrophic | Governments |
| Emergency | People | Global |
| Focus | Climate | Change |
| Started | Organisations | Field |
| Term | Threats | Action |
| Long termist | CSER | World |
| Radical | Time | Continued |

CSER = Centre for the Study of Existential Risk.
Words with the highest probability of cluster membership are at the top.
Topic names (in bold) are by default the same as the term with the highest association with the cluster.

**Table 3.** Response alternatives available to different types of futures.

| Storylines are: | Undesirable | Desirable |
| --- | --- | --- |
| Likely | Prevent beforehand, mitigate afterwards | Compare and harmonise with pre-existing plans |
| Unlikely | Monitor any possible changes in their status | Enable beforehand, exploit afterwards |

*Predefined criteria.* As previously explained, at the end of both CSER exercises, after eight iterations had been completed, participants used a widget[13] built into the ParEvo app to identify which specific surviving storylines they saw as describing the 'most likely', 'least likely', 'most desirable' and 'least desirable' futures – as seen from their own perspective. A data set of their responses was downloaded by the facilitator, then used to create a scatter plot within which there were four different quadrants of possibilities, as already shown in Figure 3 above.

As also noted earlier, a key feature of participants responses in both exercises was the diversity of participants assessments, including multiple cases of outright contradictions. Clarification and resolution of these differences in assessments is needed before there can be any agreement on what kinds of actions could be taken in response to each combination of possibilities. In exercise 1, there was some discussion of these judgements in the context of a wider-ranging group discussion about the exercise. In exercise 2, there was a more focussed but time-constrained group discussion. In future exercises, facilitators will be given additional encouragement to invest time in this stage of the ParEvo process.

Table 3 describes one aid that has already been developed to inform such discussions, but not yet used. This describes four different kinds of responses available to different types of alternative futures seen in the scatter plots. Existing theories of change (if available) would be a necessary part of the discussion of the desirable, and likely futures quadrant.

Since the completion of the CSER exercises the ParEvo app has also been adapted to enable facilitators to offer participants other customisable polar choices. For example, to describe which storylines are more/less sustainable, or more/less equitable. Combinations of these may

**Table 4.** Differences perceived between storylines, by participants in the two CSER exercises.

| Differences identified by participants in the CSER exercise 1 | Differences identified by participants in the CSER exercise 2 |
| --- | --- |
| • Substantive progress versus little progress (2). <br> • Negotiations at the international level versus regulation/implementation at the domestic level. <br> • Events are driven by exogenous shocks versus internal developments. <br> • Functional versus fractured global alliances. <br> • Reversion to earlier non-functional solutions versus new approaches. <br> • Top-down versus bottom-up political developments <br> • 'Conspiracy'-related developments versus others. | • Civil society and other actors from wider society drive change versus governments and politicians (3). <br> • Existential risk community is the main actor versus other groups also involved. <br> • More versus less agency given to individual actors. <br> • Polarised versus integrative approaches taken to existential risk. <br> • Successful/promising developments versus others less so. <br> • Global versus local impact of existential studies community (3). |

CSER = Centre for the Study of Existential Risk.

also have different implications for action. These types of analyses are likely to be relevant to development organisations pursuing sustainable development goals (SDGs).

*Emergent criteria.* Pile sorting is a method used in ethnographic enquiries (Harloff and Coxon, 2007) and in recent decades in the design of the navigation structure of websites (Olmsted-Hawala, 2006). In both CSER exercise surveys, and in many previous exercises, participants have been asked to complete an online version of a pile sorting task. Participants were asked to sort the surviving storylines into two piles according to what they thought was the 'most significant difference' between them, and then to give a text description of what that difference was. Their responses generate two kinds of data: set membership data – which storylines belong together, and qualitative data – what each set has in common and how it differs from the other. These explanations provided a participant's perspective, in contrast to the predefined differences of interest to the facilitators, embedded in the evaluation widget. Table 4 contains summaries of what were often quite detailed descriptions given by the participants. Because each of these differences are associated with a specific set of cases they could be used as attributes within a many-to-one configurational analysis of storyline outcomes.

In the first CSER exercise a more in-depth form of pile sorting was used by the authors to generate a nested classification of storylines. This was done using a method known as Hierarchical Card Sorting (HCS) (Davies, 2022). Here the same participant identifies the most significant difference between the set of storylines, then repeats the same process with each sub-set, then each sub-set, until all storylines have their own category. This generates a nested classification of storylines with a branching tree structure. Figure 6 shows one author's sort results as an example. Storyline ID numbers are provided in the leftmost column. The first distinction between the storylines was made on the adjacent column, and each column thereafter to the right describes sub-distinctions and sub-distinctions of sub-distinctions and so on. The most detailed structure (not shown here) was developed by the facilitator Tom Hobson, who had the most extensive domain knowledge.[14]

**Figure 6.** Example of a Hierarchical Card Sorting (HCS) of CSER exercise 1 surviving storylines.

While this process was cognitively demanding the resulting structure provides a good overview of the content of the surviving storylines that were generated. While it is reasonable to expect facilitators to be able to do such an analysis, it is probably less so with participants.

### Characterising the storylines as a whole

*Predefined criteria.* In both CSER exercises the exercise evaluation surveys asked participants to assess the whole set of storylines on specific predefined attributes. These included perceptions of:

- Optimism versus pessimism of the surviving storylines.
- Realism of the surviving storylines.
- Agency, in relation to the events described in the storylines.
- Influence on their thinking and behaviour at present.

These attributes were explored further using paired questions that ask first about other participant's views on the question at hand, then one's own. This provided a two-dimensional view of the range of judgements. In the second CSER exercise, scatter plots showed opinions varied widely about how optimistic the storylines were but views about realism were more in agreement.

These questions seem likely to be useful as part of a longer-term post-exercise impact evaluation, involving follow-up interviews of each participant. Degrees of optimism, realism and agency may be useful predictors of subsequent impact, among others. Text comments on expected influences could direct subsequent searches for such impacts.

*Using participant-defined attributes.* In both CSER exercises, participants were asked two complimentary questions, about content of the storylines which surprised them, and their surprise about the absence of other content from the storylines. In both exercises almost all participants did report surprising content, and slightly fewer reported surprise at what they thought was missing content (see Table 5). High levels of reporting of surprise have been found in previous exercises as well. Their prevalence supports a view of the value of collaborative development of alternative futures, though not necessarily the specific form available in ParEvo, that is, a dynamic combinatorial approach.

In future exercises, this kind of data will be subject to further analysis. One option that will be explored is the use of MSC-type questions (Davies and Dart, 2005). For example, by asking

**Table 5.** Surprises reported by participants in the two CSER exercises.

| | 1st CSER exercise | 2nd CSER exercise |
|---|---|---|
| Surprised presence | • Mention of assassinations of politicians (2)<br>• 'Conspiracy' theories re Covid (2)<br>• USA-centred nature of the storylines.<br>• Re-appearance of smallpox<br>• The role of specific institutions like the IAP<br>• A new pandemic arising from misused research, by 2023 | • The use of nuclear weapons and lack of response thereafter (2)<br>• Focus on a small number of existential risk research communities (3)<br>• Attention to dramatic impact of specific individuals (3)<br>• Disruptive technologies<br>• Bottom-up social movements |
| Surprised absence | • The collapse of the BTWC<br>• The consequences of failure to improve global biorisk management | • A major catastrophe (6)<br>• Disruptive econ/pol events<br>• Uneventful periods<br>• Religious movements<br>• Impact of EU policies/regs |

CSER = Centre for the Study of Existential Risk.

the facilitator, or the participants themselves, which of these surprises they think are the most significant of all, and why so. More work will also be done on identifying when and where contributions containing 'surprise' arise. In a more recent exercise, participants were asked to identify which specific contributions contained examples of surprise they mentioned.

*Reflection on characterisation methods.* Twelve different methods have been described above, used at three different scales of analysis. Characterisation of the whole set of surviving storylines was the least time intensive, and likely to be especially useful for post-exercise impact evaluation. Characterisation of different types of storylines seems to have a more optimal mix of effort and reward. But the diversity of participants' assessments of types of storylines does require further engagement. More detailed characterisation of individual storylines seems likely to be of most interest to individual exercise facilitators, who have the relevant domain knowledge and are likely to be willing to invest the necessary time.

If there is one lesson to be learnt from the use of the various individual methods to date it is that their use is often of most value when they are combined with each other. Automated text extraction can support manual thematic coding. Manual statistical analysis can be supported by simple machine learning methods and network visualisation. Ethnographic tools such as pile sorting can enrich facilitators' prior theories, and the development of configurational models. That said, the task of characterising long and detailed narratives in a reasonably transparent and objective manner is still a challenging one. One promising approach, only recently available, is the use of ChatGPT for the analysis of narratives (Davies, 2023).

## Assessing the impact of ParEvo exercises

The expected post-exercise impacts of a ParEvo exercise vary from exercise to exercise, depending on the individual facilitator's objectives. These have included:

- Influencing the content of a strategic plan (1 exercise completed, 1 planned).
- Lead to the publication of papers in an academic journal (2 exercises).
- Inform the content of an evaluation (4 exercises).

- Lead to the revision of risk management protocol (1 exercise).
- Changes to plans for ensuring research uptake (1 exercise).

There are constraints on the extent to which the impact of these exercises can be identified by the platform administrator. Facilitators are not obliged to share exercise evaluation data, or information about the subsequent effects of their exercise. Publication of exercise data generated by the app itself can only be used by the administrator with the consent of the facilitator and is subject to the constraints of the privacy policy.[15] However, the administrator does encourage facilitators to publish their own papers on the design and results of their own exercises.

At the platform level, objectives are not exercise specific. These have included:

1. The development of new features in the ParEvo app. This was especially the case with the six earliest exercises facilitated by the administrator. And has continued as a secondary objective thereafter. Visible improvements have included the development of a comment facility, a tagging facility, a leaderboard and more flexible evaluation options.
2. Further exercises by first-time facilitators. This has been the case with four facilitators, leading to six additional exercises to date.
3. Requests by other organisations wanting to use ParEvo for the first time. Four new organisations were registered as users in 2022, with exercises planned in the next 12 months.
4. Accumulation of data from multiple exercises that is sufficient to enable analyses of exercise parameters and how they affect exercise outcomes. This article reflects early developments towards this objective.
5. Publication of papers about ParEvo in academic journals and books. Five are in process to date.
6. Recovery of investment costs, through payments for technical support provided to organisations using ParEvo. This has been underway since early 2022, although pro bono technical support for first time users remains the norm, as does free use of the app itself.

Objectives for the platform have changed over time. With the earliest exercises the main objective was to ensure that the web application functioned as expected. Then more attention was given to the development of evaluation options, within the app itself, and using third-party surveys, for example, SurveyMonkey. In the last 12 months, more emphasis has been given to ensuring sufficient post-exercise facilitated discussion among the participants, to work through the implications of the exercise. This needs to continue. More encouragement is also being given to exercise facilitators to articulate their objectives for their exercises before they start. Both facilitators did so in the two recent CSER exercises.

On a more continuous basis, it has always been intended that knowledge accumulated from exercises completed to date will be distilled and documented in the supporting website available to all users, and others (ParEvo, 2022). This website now includes a framework for thinking about the implications of different alternative futures, as classified using the criteria of desirability and likelihood, shown in Table 3. There is also extensive guidance on the analysis of participants' contribution behaviour, using simple network analysis tools, informed by a diversity perspective. The development of generically useful coding frameworks is also under

consideration, given the prevalence to date of some common features across exercises. For example, the frequent reference to the emergence and dissolution of different types of actor coalitions, and conflicts between these.

## Alternative futures for ParEvo

While all six of the platform objectives listed above will remain relevant, some additional features of these are now in sight. One is the development and testing of individual performance measures that could introduce an element of gamification to how people participate in exercises. Another is to develop new means by which participation in exercises could be scaled up, well beyond the 10 to 15 participants per exercise seen so far. Seven options have already been identified, involving different forms of involvement (i.e. commentators and observers), and a variety of ways of involving the participating contributors themselves.[16] In terms of new users and new uses of ParEvo, it is hoped that this introduction to ParEvo will help widen how evaluators think about the future and help others to do so, especially when it comes to developing and updating theories of change. More ambitiously, it may be feasible to engage current and previous facilitators as participants in a ParEvo exercise about the future uses of ParEvo itself.

## The future relationship between ParEvo exercises and theories of change

Theories of change, as seen by evaluators, have a particular perspective on the future. They are characteristically:

- Predictive, of what kinds of future will happen, given specific conditions.
- Focused specifically on a desirable and likely future.
- Convergent, involving a combination of various activities that lead to that future.
- Intended to assist a planned response that will lead to that future.

In terms introduced by March (1991), and relative to ParEvo, they are largely about the 'exploitation' of existing knowledge, as district from 'exploration' of the new. In contrast, ParEvo exercises are much more about exploration. Their futures are not predictive, instead they describe a range of possibilities. These can involve various combinations of desirability and likelihood, and other attributes. They are divergent, describing multiple alternative developments that might arise from a single source event. And they are intended to facilitate adaptive thinking about how to respond, proactively and retroactively, to a range of possible futures.

As emphasised by March, and others after him (Wilden et al., 2018), neither approach is intrinsically superior to the other. Both are useful in different circumstances. Having the capacity to both is better than only having one. Making use of applications like ParEvo is one way of doing so.

**Declaration of conflicting interests**

## ORCID iDs

Rick Davies [iD] https://orcid.org/0000-0002-6889-3314

Tom Hobson [iD] https://orcid.org/0000-0002-1244-3787

Lara Mani [iD] https://orcid.org/0000-0003-2967-7499

Simon Beard [iD] https://orcid.org/0000-0002-2834-0993

## Notes

1. Search for 'Africa' + 'closed projects', average duration of first 27 finds with both start and end dates provided.
2. Allegedly by the Prussian Field Marshal Helmuth von Moltke the Elder in 1871 https://quote investigator.com/?s=No+battle+plan+ever+survives+contact+with+the+enemy
3. Futures studies or Futures thinking refers to systematic approaches to thinking about the future. 'Foresight' refers to the specific act of applying futures tools, processes or methods.
4. A full list of completed and planned exercises is maintained here: https://mscinnovations.word-press.com/list-of-parevo-exercises/
5. Referred to hereafter as 'exercise 1'.
6. Referred to hereafter as 'exercise 2'.
7. But less so if there are not enough iterations for all these possibilities to occur, and if some participants are absent from an iteration.
8. Balance can also be calculated by examining the number of instances of each type of combination.
9. Here their scale position is a net value. The number of participants selecting a storyline as least likely/least desirable was subtracted from the number who selected the same storylines as likely/desirable.
10. Or alternately, most likely, and most desirable, and least likely and least desirable.
11. This option is explored in detail at https://mscinnovations.wordpress.com/alt-versions/#ToC
12. By itself, having a classification accuracy of 72% (12 true positives, 5 false positives, 14 false negatives, 37 true negatives.
13. A widget is a component that is added to a website or application as a stand-alone feature,
14. Not shown here because of insufficient room for display of the detailed text.
15. The full text of the privacy policy is available here https://parevo.org/privacy/
16. See https://mscinnovations.wordpress.com/alt-versions/#scaled-up-parevo

## References

Austrian Development Agency (ADA) (2022) *Evaluability Assessments in Austrian development Cooperation: Guidance Document*. Available at: https://www.entwicklung.at/fileadmin/user_upload/Dokumente/Evaluierung/GL_for_Evaluability_Assessments.pdf (accessed 11 August 2022).

Avin S, Gruetzemacher R and Fox J (2020) Exploring AI futures through role play. In: *Proceedings of the AAAI/ACM conference on AI, ethics, and society*, New York, 7–9 February, 8–14. New York: ACM.

Barbrook-Johnson P and Penn AS (2022) *Systems Mapping: How to Build and Use Causal Models of Systems*. Cham: Palgrave Macmillan.

Beard SJ, Cooke N, Dryhurst S, et al. (2023) Exploring Futures for the Science of Global Risk. 4405991, SSRN Scholarly Paper. Rochester, NY. Available at: https://papers.ssrn.com/abstract=4405991 (accessed 31 March 2023).

Bengston DN (2016) The futures wheel: A method for exploring the implications of social–Ecological change. *Society & Natural Resources* 29(3): 374–9.

Borgatti SP, Everett MG and Freeman LC (2002) *UCINET for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Borgatti SP, Everett MG and Johnson JC (2018) *Analyzing Social Networks*. Los Angeles, CA: SAGE.

Boyd-Graber J, Hu Y and Mimno D (2017) Applications of topic models. *Foundations and Trends in Information Retrieval* 10(10): 158.

Campbell CM, Izquierdo EJ and Goldstone RL (2022) Partial copying and the role of diversity in social learning performance. *Collective Intelligence*. Epub ahead of print 13 September. DOI: 10.1177/26339137221081849.

Cassidy M and Mani L (2022) Huge volcanic eruptions: Time to prepare. *Nature* 608(7923): 469–71.

Christian B (2020) *The Alignment Problem: Machine Learning and Human Values*, 1st edn. New York: W. W. Norton & Company.

Clark H (2022) TOCO. Available at: https://www.actknowledge.org/toco-software/ (accessed 30 May 2022).

Centre for the Study of Existential Risk (CSER) (2022) Available at: https://www.cser.ac.uk/ (accessed 19 August 2022).

Davies R (1998) *Order and diversity: Representing and assisting organisational learning in non-government aid organisations*. PhD Thesis, University of Wales Swansea, Swansea.

Davies R (2018) *Representing Theories of Change: Technical Challenges with Evaluation consequences*. CEDIL. Available at: https://cedilprogramme.org/wp-content/uploads/2018/11/Inception-Paper-No-15-Rick-Davies-Representing-theories-of-change.pdf (accessed 21 August 2022).

Davies R (2022) Hierarchical Card Sorting (HCS). *Monitoring and Evaluation NEWS*. Available at: https://mande.co.uk/special-issues/hierarchical-card-sorting-hcs/ (accessed 19 August 2022).

Davies R (2023) Using ChatGPT as a tool for the analysis of text data. *Monitoring and Evaluation NEWS*. Available at: https://mande.co.uk/2023/lists/software-lists/using-chatgpt-as-a-tool-for-the-analysis-of-text-data/ (accessed 10 February 2023).

Davies R and Dart J (2005) *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. Available at: https://www.researchgate.net/publication/275409002_The_'Most_Significant_Change'_MSC_Technique_A_Guide_to_Its_Use (accessed 21 August 2022).

Dennett DC (1996) *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. London: Penguin Books.

Edelman GM (1987) *Neural Darwinism – The Theory of Neuronal Group Selection*. New York: Basic Books.

EvalC3 (2022) About EvalC3. Available at: https://evalc3.net/ (accessed 19 August 2022).

Feinstein O (2019) *Checklist for Evaluation Recommendations*. The Evaluation Center, Western Michegan University. Available at: https://wmich.edu/sites/default/files/attachments/u3036/2019/eval-rec-feinstein.pdf (accessed 19 August 2022).

Funnell SC and Rogers PJ (2011) *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Hoboken, NJ: John Wiley & Sons.

Harloff J and Coxon APM (2007) *How To Sort: A Short Guide on Sorting Investigations*. Available at: http://methodofsorting.com/HowToSort1-1_english.pdf (accessed 24 October 2012).

Hobson T, Edwards B, Sutherland W, et al. (2022) Submission of evidence to the cabinet office enquiry on the biological security strategy. *Report, Cabinet Office, London*, 30 March.

Kemp L, Xu C, Depledge J, et al. (2022) Climate Endgame: Exploring catastrophic climate change scenarios. *Proceedings of the National Academy of Sciences* 119(34): e2108146119.

Maas M, Cooke D, Hobson T, et al. (2021) Reconfiguring resilience for existential risk: Submission of evidence to the Cabinet Office on the new UK National Resilience strategy. *Report, Centre for the Study of Existential Risk (CSER), University of Cambridge, Cambridge*, 27 September.

March JG (1991) Exploration and exploitation in organizational learning. *Organization Science* 2(1): 71–87.

Morell JA (2022) Complexity-informed evaluation – An exploration in understanding pattern, predictability, and how change happens. In: Evaluation Uncertainty. Available at: https://evaluation uncertainty.com/2022/04/01/complexity-informed-evaluation-an-exploration-in-understanding-pattern-predictability-and-how-change-happens/ (accessed 18 February 2023).

Olmsted-Hawala E (2006) Card sorting, information architecture and usability: Adding in our users' perspective to re-design the Census Bureau Web Site. In: *Proceedings of the 2006 IEEE international professional communication conference*, Saratoga Springs, NY, 23–25 October, 27–33. New York: IEEE.

Organisation for Economic Co-operation and Development (OECD) (n.d.) *Futures & Foresight*. Observatory of Public Sector Innovation. Available at: https://oecd-opsi.org/guide/futures-and-foresight/ (accessed 17 February 2023).

Page SE (2008) *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ: Princeton University Press.

Page SE, Lewis E, Cantor N, et al. (2017) *The Diversity Bonus: How Great Teams Pay off in the Knowledge Economy*. Princeton, NJ: Princeton University Press.

ParEvo (2022) Introduction to ParEvo. Available at: https://mscinnovations.wordpress.com/ (accessed 19 August 2022).

Raford N (2015) Online foresight platforms: Evidence for their impact on scenario planning & strategic foresight. *Technological Forecasting and Social Change* 97: 65–76.

Rhydderch A (2017) *Scenario Building: The 2×2 Matrix Technique*. Available at https://www.researchgate.net/publication/331564544_Scenario_Building_The_2x2_Matrix_Technique (accessed 12 August 2022).

Rogers P (2014) *Theory of change*. Methodological briefs impact evaluation no. 2. UNICEF. Available at: https://www.betterevaluation.org/sites/default/files/Theory_of_Change_ENG.pdf (accessed 12 August 2022).

Rogers P and Macfarlan A (2020) *What Is Adaptive Management and How Does It Work?* Department of Foreign Affairs and Trade (DFAT). Available at: https://www.betterevaluation.org/sites/default/files/2021-11/MandE_for_Adaptive_Management_WP2_What_Is_AM_%26_how_does_it_work_202009.pdf (accessed 12 August 2022).

Simon HA (1956) Rational choice and the structure of the environment. *Psychological Review* 63: 129–38.

Starr L and Peek N (2019) *Theory of Change Review as an Adaptive Management Tool*. USAID Learning Lab. Available at: https://usaidlearninglab.org/community/blog/theory-change-review-adaptive-management-tool (accessed 15 August 2022).

Stirling A (2007) A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface* 4(15): 707–19.

Text2Data.com (2022) Available at: https://text2data.com (accessed 19 August 2022).

The BigML Team (2022) *Topic Modeling with the BigML Dashboard*. BigML. Available at: https://static.bigml.com/static/pdf/BigML_Topic_Modeling.pdf?ver=02482c9 (accessed 21 August 2022).

United Nations Development Program (2018) *Foresight Manual: Empowered Futures for the 2030 Agenda*. UNDP Global Centre for Public Service Excellence. Available at: https://www.undp.org/sites/g/files/zskgke326/files/publications/UNDP_ForesightManual_2018.pdf (accessed 17 February 2023).

Waverley Consultants (2017) *The Futures Toolkit: Tools for Futures Thinking and Foresight Across UK Government*. Government Office for Science. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674209/futures-toolkit-edition-1.pdf (accessed 17 February 2023).

Wilden R, Hohberger J, Devinney TM, et al. (2018) Revisiting James March (1991): Whither exploration and exploitation? *Strategic Organization* 16(3): 352–69.

Witze A (2020) How a small nuclear war would transform the entire planet. *Nature* 579(7800): 485–7.

World Futures Studies Federation (WFSF) (n.d.) Futures Publications – Journals. Available at: https://wfsf.org/futures-publications-journals/ (accessed 20 August 2022).

Xiao T, Makhija M and Karim S (2022) A knowledge recombination perspective of innovation: Review and new research directions. *Journal of Management* 48(6): 1724–77.

Dr. Rick Davies is an independent evaluation consultant and developer of evaluation tools. He has worked with NGOs, bilateral and multilateral institutions on the evaluation of international development aid programmes.

Tom Hobson is a CSER Research Associate. Tom's policy-oriented work concerns the governance of emerging technologies, biotechnology and biosecurity. His research is on how expert communities and policy actors imagine the future.

Lara Mani is a CSER Research Associate. Her research explores the efficacy of various communication methods and strategies to address the mitigation and prevention of global catastrophic risks (GCRs)

Simon Beard is the Academic Programme Manager and Senior Research Associate at CSER. He works across CSER's various research projects, covering issues of ethics, research methods, decision making and communications.