WILEY

View PDF Redline PDF

Article DOI : 10.1002/ev.20606

Author Proof

CHAPTER

# Integrating foresight and evaluation: The example of ParEvo

Rick J. Davies [1]

AQ1

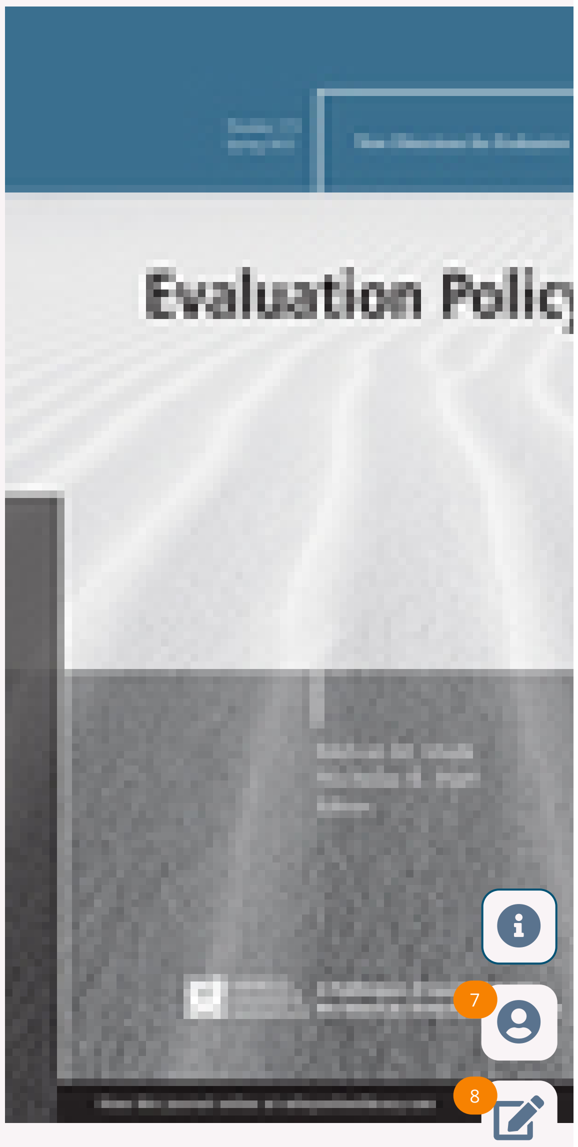
[1]Independent Evaluation Consultant, Cambridge, UK

**Correspondence**
Rick J. Davies, Independent Evaluation Consultant, Cambridge, UK.
Email: rick.davies@gmail.com



## Abstract

This paper aims to encourage connections between the fields of program evaluation and foresight practice by describing the use of a web application ParEvo as an example of how foresight methods can incorporate evaluative thinking. It begins by introducing the field of foresight practice, along with its differences and commonalities with that of evaluation. It then introduces ParEvo.org, and its uses to date, along with a detailed description of how a ParEvo exercise is implemented. An overview is also given of how different types of evaluation activities are built throughout the implementation process. Attention is then focused on three different types of embedded theories of change, and how they can be evaluated. The paper concludes with two observations about the breadth and depth of evaluative thinking present in ParEvo processes and the supporting platform. Some limitations are also acknowledged, arising from the rich nature of the outputs generated and the decentralized nature of the design and operation of ParEvo exercises.




1 Proof Initiated
7/17/2024

2 Corresponding Author
Due date: 7/19/2024
Start date: 7/17/2024
End date: 7/19/2024

3 Proof Collator
Due date: 7/21/2024
Start date: 7/19/2024

# INTRODUCTION

AQ2

## What is foresight and how does it relate to evaluation?

Foresight, as a term used in the field of futures studies, “refers to the capacity to think ahead, consider, model, create, and respond to future possibilities” (Hejazi, 2023). It is a universal human capability to proactively engage with the future, taking into account values, expectations, choices, and responsibilities. Foresight approaches vary in their focus from the near to distant futures. Some are very exploratory, seeking to identify a wide range of potential future scenarios and developments, often through techniques such as trend analysis and horizon scanning (Dart & Gates, 2024; Waverley Consultants, 2017). Some are more normative, using methods to envision preferable futures, and considering ethical, social, and environmental values. These include approaches like backcasting, which starts with a desirable future and work backward to identify the steps needed to reach it (Wikipedia, 2024). Some focus on anticipatory responses, on preparing for potential future disruptions, risks, and uncertainties. They include techniques such as risk assessment, wild card analysis, and resilience planning (Conrad et al., 2023). Some prioritize engage a wide range of stakeholders in the foresight process, such as citizens, experts, and decision-makers, to ensure diverse perspectives are considered in shaping the future (Hamel et al., 2021). Foresight methods also vary in their emphasis on quantitative versus qualitative methods. The former uses data, models, and statistical analysis to forecast future trends. The latter uses expert opinions, scenario building, and narrative storytelling to explore future possibilities and uncertainties. For a structured view of this diversity see Poli's (2018) classification of future-related methods.

Within this substantial diversity, there are some common threads. In the face of the inherent uncertainty of the future, these methods aim to provide a systematic and structured way of thinking about the future. Many foresight methods emphasize the importance of involving a wide range of stakeholders in the process, such as participatory foresight. The development of alternative scenarios is a widespread practice (Bishop et al., 2007). These are narrative or analytical descriptions of possible, probable, and preferable futures, designed to explore different outcomes that could result from current trends, wild cards, uncertainties, and strategic decisions. Foresight methods often involve identifying and analyzing trends, drivers, and uncertainties that could shape the future. These include technological, economic, environmental, political, and social factors. These methods acknowledge the unpredictability of the future and therefore focus on flexibility and adaptability. They often aim to prepare organizations or individuals for a range of possible outcomes rather than predicting a single future (Popper, 2010).

In this respect, foresight thinking and methods differ significantly from the way many evaluators think and work. Evaluations are primarily concerned with assessing current or past programs, projects, or policies, not the future. They look backward or at the present to understand the effectiveness or impact of an initiative. Evaluations typically (but not exclusively) involve measuring performance against predefined objectives or criteria. They rely heavily on verifiable data and evidence to assess outcomes and impacts. And they are generally more focused and specific, targeting a particular program, project, or policy. In contrast to foresight, evaluations will often be linked to a singular theory of change, representing a specific view of the future, one that is both desirable and likely. They are in effect predictions, albeit often surprisingly optimistic ones. While a theory of change might include a wide range of activities, these are expected to be convergent, contributing to the achievement of an expected and desired outcome. Not surprisingly, good arguments have recently been made for evaluators to take a wider perspective on the future (Carden, 2023).

Using the terms first introduced by the organizational learning theorist James March (1991), theories of change, and their subsequent evaluation are largely about the “exploitation” of existing knowledge (i.e., making full use of existing resources). As distinct from “exploration” of the new. In comparison, foresight activities are much more about “exploration.” As emphasized by March, and others after him (e.g., Wilden et al., 2018), neither of these two approaches is intrinsically superior to the other. Both are useful in different circumstances and having the capacity to do both is better than only having one.

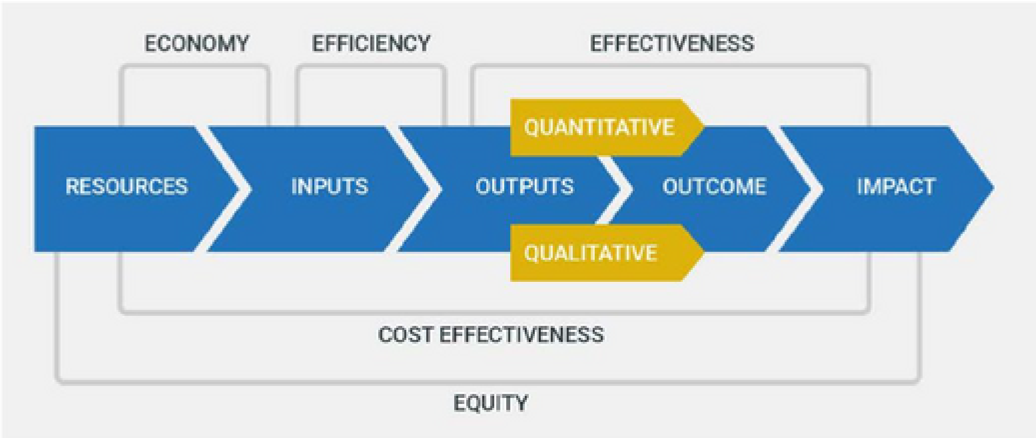
## Is foresight evaluable?

Given the above description of foresight, it might be reasonable to think that views of the future are intrinsically unevaluable. The events described have not yet happened and may never happen. Even if they do happen, the future events are so distant that it is not practical to assess the value of a foresight exercise taken many years earlier. This argument is faulty in two different ways, depending on the way it is phrased. Firstly, if the argument assumes that foresight is about prediction, then the testing of a long-term prediction, especially involving complex social changes, can readily be acknowledged as inherently problematic. But this assumption does not fit the ambitions of most of those involved in

WILEY

View PDF Redline PDF

foresight exercises, who are not aiming to make testable predictions. Secondly, if the ambitions of foresight exercise are couched in terms of improving preparedness for the future, rather than making predictions, again there may be valid concerns about practicability. But this only applies if evaluation efforts are focused on “final” outcomes, that is, the averting of a final crisis, or the realization of long-term opportunities. However, this view is too narrow. In the field of evaluation, it is well recognized that the focus of an evaluation can span anywhere from the provision of inputs into an intervention (including foresight exercises), to the outputs generated, to their near and long-term outcomes, and their impacts. That range is captured in a DFID (2011) publication on “value for money” (broadly interpreted), as seen in Figure 1. Evaluation at these earlier stages can be both formative and summative in its effects. The use of findings can prove the immediate effectiveness of foresight activities. And serious failures identified in the earlier stages are potentially valid predictors of the absence of longer-term impacts.




AQ3

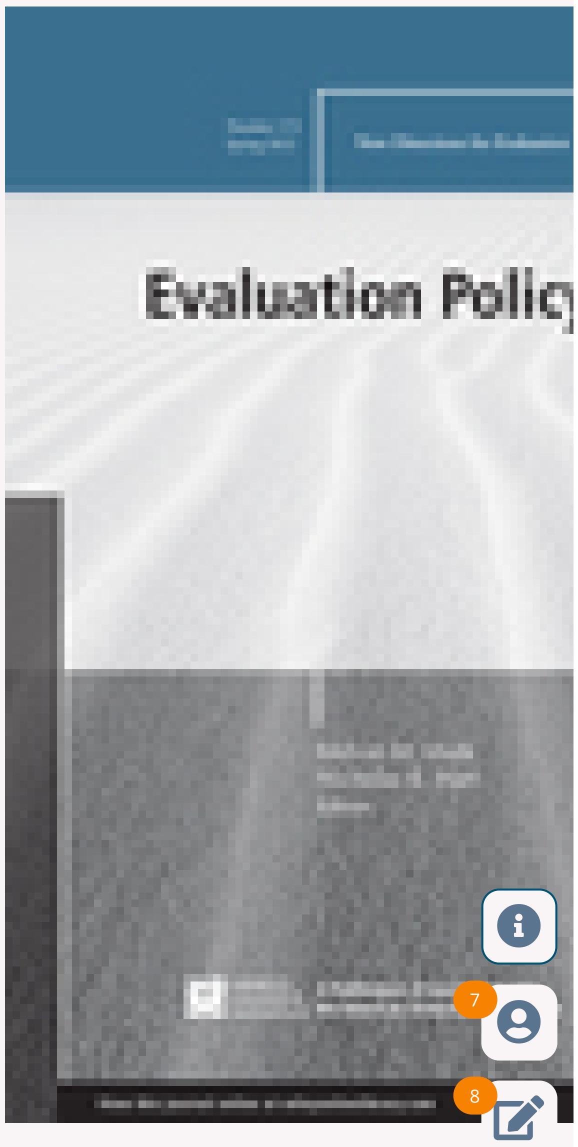
**FIGURE 1.**
The potential range of evaluation concerns.

In sum, the fields of evaluation and foresight cover not only two ends of a temporal spectrum, ranging from the past, through the present and into the future. But they also overlap and these activities can be combined. Evaluation activities can be built into foresight practices. And as shown elsewhere in this issue (e.g., Blundo-Canto et al., 2024; Dart & Gates, 2024; Halonen et al., 2024) foresight can be built into evaluation activities. The purpose of this article is to show by example how they can be creatively combined. This will be done by discussing the design and use of the web application ParEvo (Davies et al., 2023).

While this article is primarily directed to an audience of evaluators, it is also expected to be of interest to foresight practitioners who want to incorporate evaluative processes into their activities. With this in mind, it is expected that this paper will also be shared in foresight networks, such as the Association of Professional Futurists, the World Futures Studies Federation , and the World Future Society.

## WHAT IS ParEvo?

ParEvo is an online method of exploring alternative futures or histories, using a participatory evolutionary process (hence ParEvo) (Davies, 2019). The process is designed to be used by multiple people to produce a branching structure of storylines about what could happen. ParEvo was developed by an evaluation consultant and had its origins in research on organizational learning (Davies, 1998:337). Its design consciously includes many forms of evaluative activity. But at the same time, its outputs are intended to widen participants’ views of the future well beyond the ambit of a single theory of change. ParEvo is a process that is available to multiple users via a dedicated site, typically described as a platform (McAfee & Brynjolfsson, 2017).



1 Proof Initiated
7/17/2024

2 Corresponding Author
**Due date:** 7/19/2024
**Start date:** 7/17/2024
**End date:** 7/19/2024

3 Proof Collator
**Due date:** 7/21/2024
**Start date:** 7/19/2024

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 theories of change (Clark, 2022), these fit the description above: singular, and convergent on desired and likely changes. They do not focus on the development of a diversity of views about possible futures. Other online platforms have been developed by futurists for the exploration of the future, some of which have been reviewed by Raford (2010). ParEvo exercises also 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 ParEvo process is more divergent and satisficing rather than convergent and optimizing—as is the case with online Delphi exercises, perhaps the most widely known foresight method. Third, with ParEvo the construction of narratives precedes and informs a detailed analysis, rather than following 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, it generates a greater number of alternative futures when compared to the average futures exercise, perhaps most stereotypically represented in a 2 × 2 matrix (Rhydderch, 2017). Finally, there is the greater prominence that it gives to evaluation activities within the whole process.

As of late 2023, 19 different ParEvo exercises have been completed during and since the development of the app (Davies, 2020). Participants have included school students, volunteers from evaluation communities of practice, crowdsourced adult university-educated UK participants, UK think tank staff, UN Volunteers, UN agency staff members, international civil society organizations representatives, and internationally recognized experts in particular scientific fields. Futures explored include post-Brexit Britain, post-Trump USA, corporate strategies, the future of international civil society (Hutchings & Mugenzi, 2023), uptake pathways for education research, future global climate changes, the future of existential risk studies (Beard et al., 2023), and the governance of experiments in solar radiation management (Futerman & Beard, 2023). Alternative histories have also been explored, capturing different versions of what actually happened. Covering topics like agricultural development projects, UNV volunteer experiences, and UN agency gender policy implementation. Eight of the earlier exercises were initiated by the lead author, and 11 of the more recent exercises were initiated and facilitated by others, including researchers, consultants, and staff members.

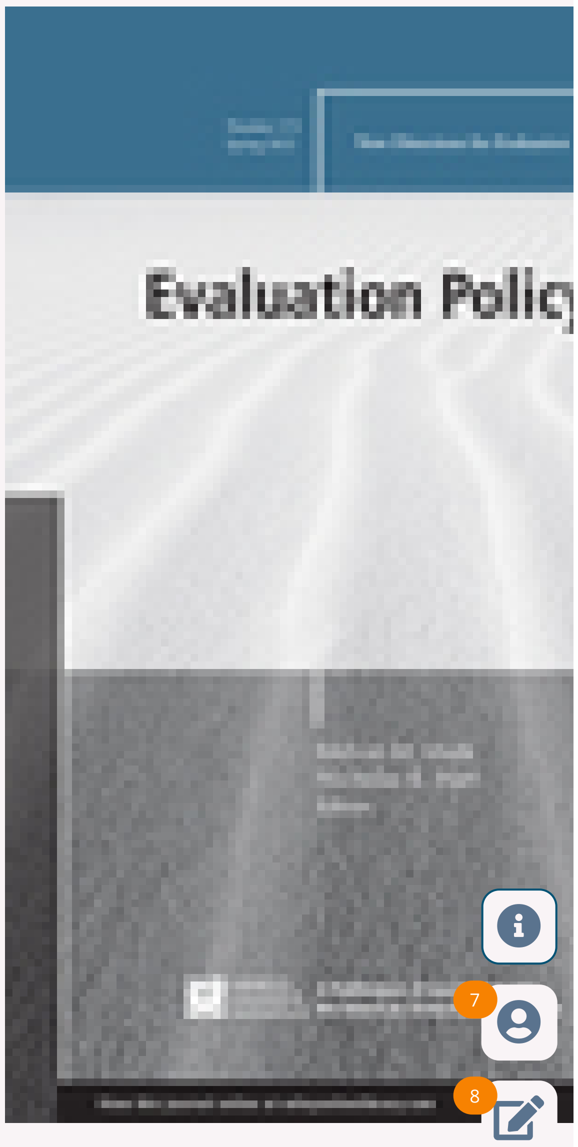
## HOW ParEvo WORKS

In summary, ParEvo is a method of developing alternative futures (or past histories), using a participatory version of the evolutionary algorithm (Campbell, 1960; Dennett, 1996). This involves the re-iteration of variation, selection, and reproduction of ideas about how future events will unfold. The process is designed to be used by multiple people, to produce a collective good: a branching set of storylines about different possible futures. In addition, the process generates data on the structure of participation—how people have interacted to produce those storylines.

### The generation of storylines

Via an online interface at the ParEvo.org platform, exercise participants are presented with a seed text equivalent to the first paragraph of a novel, which is prepared by an exercise facilitator.[1] Participants are then 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 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 more participants. As this process is reiterated what emerges is a tree structure of alternative storylines (Figure 2).

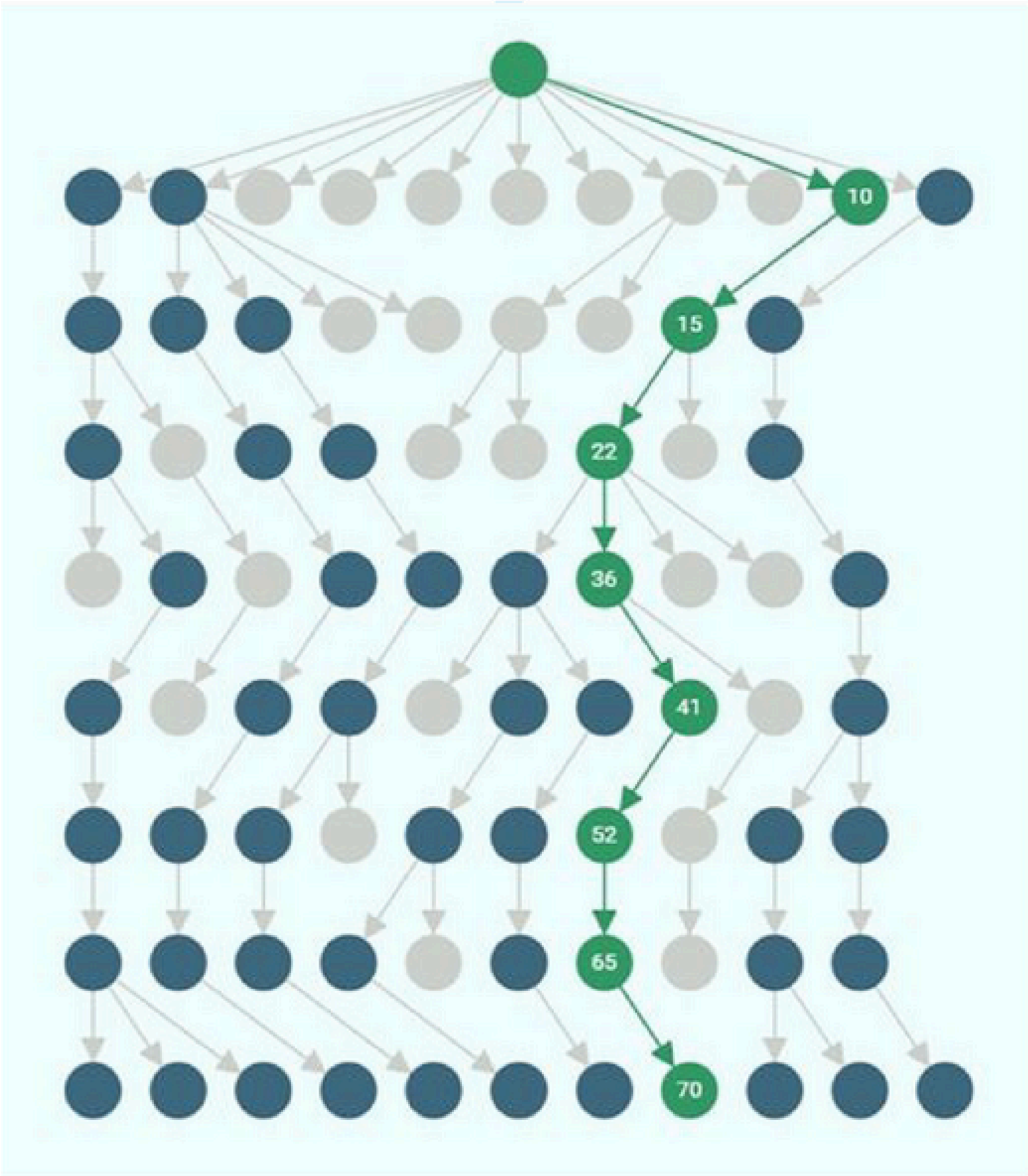


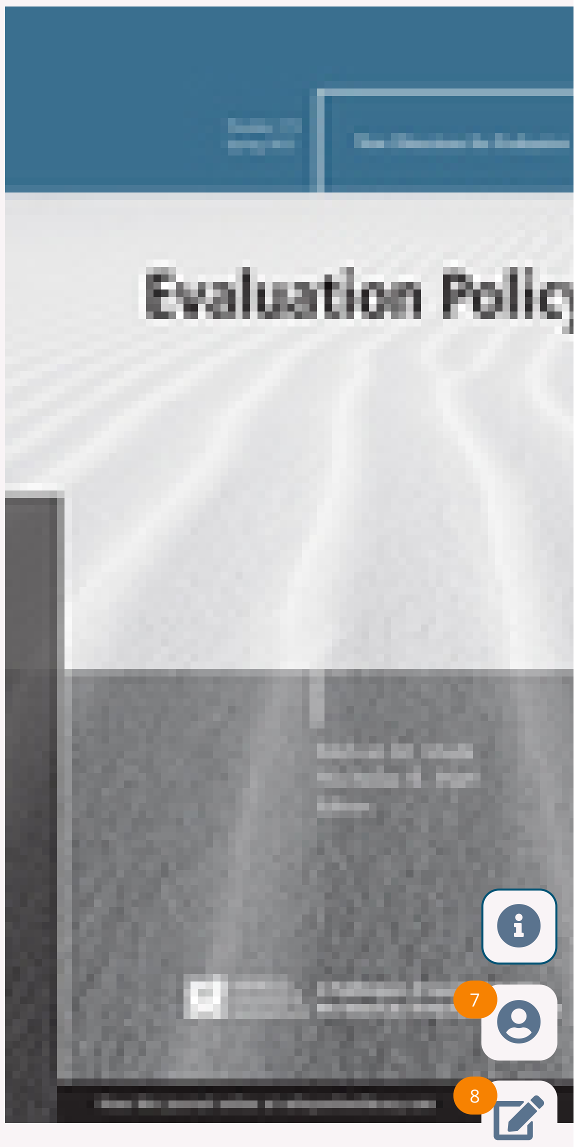

**FIGURE 2.**
Tree structure of alternative storylines generated by the Center for the Study of Existential Risk (CSER) (Davies et al., 2023).

In the Figure 2 example 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. Gray nodes in the tree structure indicate paragraphs which were not continued and which in effect represent extinct storylines. Dark gray nodes represent paragraphs which others did build on and which became part of storylines that survived until the end of the exercise (bottom row). Numbered nodes represent one storyline which has been highlighted by the cursor of 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).

## The evaluation of storylines

The evaluation of storylines has two general purposes, concerning process and contents. Both facilitators and the administrator typically want to understand how participants viewed the *process* and how it could be improved. Understanding the storyline *contents* is relevant to two kinds of overall objectives: cognitive and behavioral. That is, being

View PDF Redline PDF




1 Proof Initiated
7/17/2024

2 Corresponding Author
**Due date:** 7/19/2024
**Start date:** 7/17/2024
**End date:** 7/19/2024

3 Proof Collator
**Due date:** 7/21/2024
**Start date:** 7/19/2024

more critically aware of how we think about the future and being better prepared for those futures through actions that can be planned in advance.

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 (e.g., Davies, 2022). Both allow participants to make polar evaluation choices of which storyline they think is most or least likely, desirable, equitable, sustainable, and any other polarities as needed. The available choices are set by the exercise facilitator. The online surveys used to date have also used both open- and closed-ended questions about the storyline contents and exercise process. Facilitators can access a list of questions covered by previous surveys on a supporting website (Davies, 2022). In addition, the facilitator can download from ParEvo.org multiple Excel datasets containing automatically generated information about the contents of the storylines and the structure of people's interactions in the construction of the storylines. Exercise facilitators have also organized 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.

In addition to the basic options as described, there are additional features providing other evaluative opportunities. Participants can be enabled to use a Comment facility, after each iteration, to ask questions or make comments on the most recent contributions of any of the participants. This facility can also be made available to third parties, such as technical specialists. The whole exercise can also be made selectively viewable by a defined group of "observers," whose views can then be sought by inclusion in any online survey and through direct contact by the facilitator. A tagging facility is also available, allowing participants to describe their contributions, with choices from a menu of labels predefined by the facilitator.

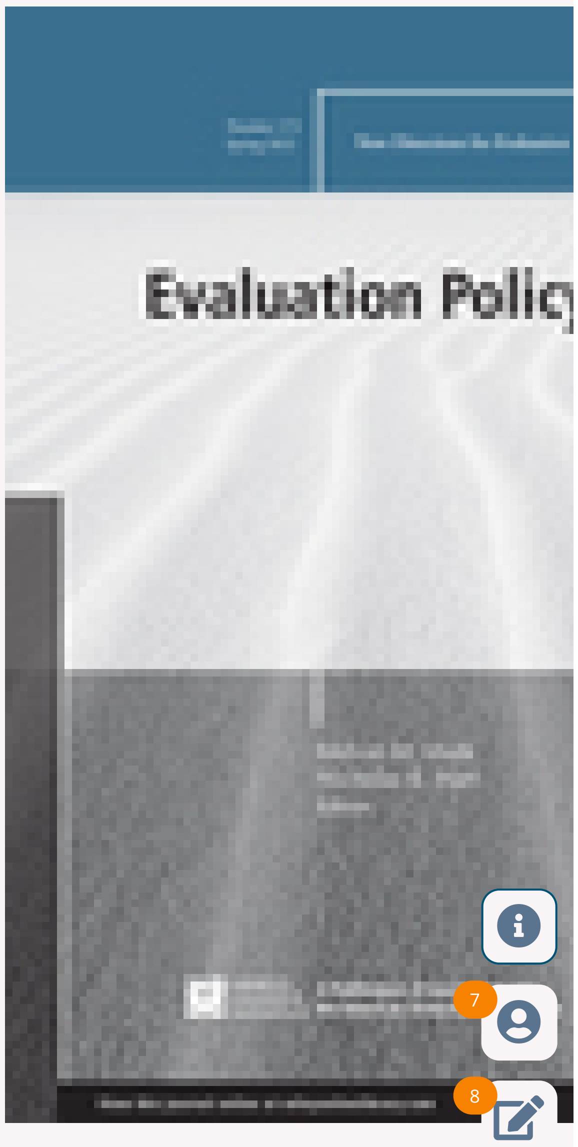
## THEORIES OF CHANGE WITHIN ParEvo

The identification and analysis of the Theory of Change is as appropriate to the evaluation of a foresight exercise, such as ParEvo, as it is to many other interventions designed to improve human welfare. There are three different types of theories of change about what happens with ParEvo exercises: (a) storyline level theories, (b) exercise level theories, and (c) platform level theories. Within each of these, the focus can vary at different points of time, of the kind described in Figure 1.

### Storyline level theories

The first kind of theory of change is the participants' own often informal and tacit theories about what might happen in the future, as evident in the contents of their text contributions and the resulting composite storylines. A storyline typically describes a sequence of actions over time, causally linked to each other, involving different actors and institutions, believing in different values and pursuing different objectives, in specific contexts. Examples of such storylines can be seen online in the exercise completed by the International Centre for Civil Society (Hutchings & Mugenzi, 2023). These narrative accounts are much more multidimensional than the skeletal descriptions provided by diagrammatic theories of change that many evaluators are familiar with. Embedded in those text descriptions are tacit and explicit assumptions and beliefs, which can be subject to scrutiny. As noted above, one of the two types of objectives of a ParEvo exercise is cognitive: to identify and address gaps and weaknesses in how we think about the future.

View PDF Redline PDF




1 Proof Initiated
7/17/2024

2 Corresponding Author
**Due date:** 7/19/2024
**Start date:** 7/17/2024
**End date:** 7/19/2024

3 Proof Collator
**Due date:** 7/21/2024
**Start date:** 7/19/2024

The main focus of storyline evaluation to date has been on characterization and identifying important points of difference between storylines. Differences of interest have included storylines that have survived versus gone extinct; storylines at different stages of development during the exercise, and storylines created by multiple versus few participants. One important means for doing so has been the use of what ethnographers call pile or card sorting questions which ask participants what they see as the most significant difference between one group of storylines and the other.

Another method of analysis has been to analyze the distribution of kinds of storylines, as described by pre-defined criteria, such as likelihood and desirability, mentioned above. Two example scatter plots from two exercises carried out by the Centre for the Study of Existential Risk are shown in Figure 3. Three features are of interest. Firstly, if there are any empty quadrants that represent omitted futures[2]. In effect, following the advice by Woo (2017), that "In order to think out of the box, you first need to find the box." The second are storylines which have received contradictory ratings by different participants (shown in darker gray), and which need resolution. The third are the storylines which are unambiguously in one quadrant or the other. Here the next question is what can or should be done about these types of futures. What can be done to avert or enable them before they happen, and to mitigate or exploit them if they do happen? These questions then need to be addressed in stakeholder workshops of the kind that followed the ParEvo workshops, and which are now routinely held after the evaluation stage.

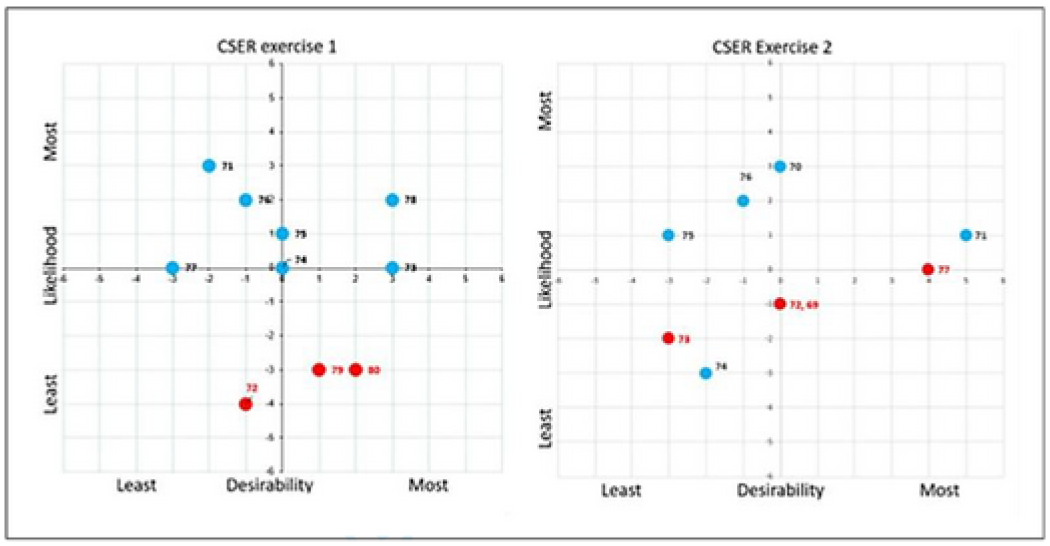



**FIGURE 3.**
Distribution of evaluations of the surviving storylines, in two CSER exercises. Numbered data points refer to individual storyline IDs. The darker data points indicate storylines that received contradictory ratings.

Exercise facilitators have also been active evaluators of the contents of the generated storylines. A wide range of content analysis methods have been applied to ParEvo storylines, as summarized recently by the author (Davies, 2023). These involve thematic coding of storyline contents using a schema that are specifically relevant to the types of futures being explored. Facilitators can apply their own analytic framework and concerns, and address questions that are applicable across the whole set of storylines.

The general purpose of the analysis of storylines is to gain greater awareness of how people are thinking about the future and how that might be limiting their capacity to prepare for futures, both known and unknown.

## Exercise level theories

The second type of theory of change is the exercise facilitator's expectations of what they want to see happen with the exercise they have designed and organized, as evident both in their choice of the design parameters of the exercise and in the kinds of expected outcomes that are of interest. The design parameters include the kind of futures they want to see explored, the kinds of people to be involved, the time span and granularity of the exercise iterations, and the guidance they give to participants at the beginning of each new iteration.

Expected outcomes have included:

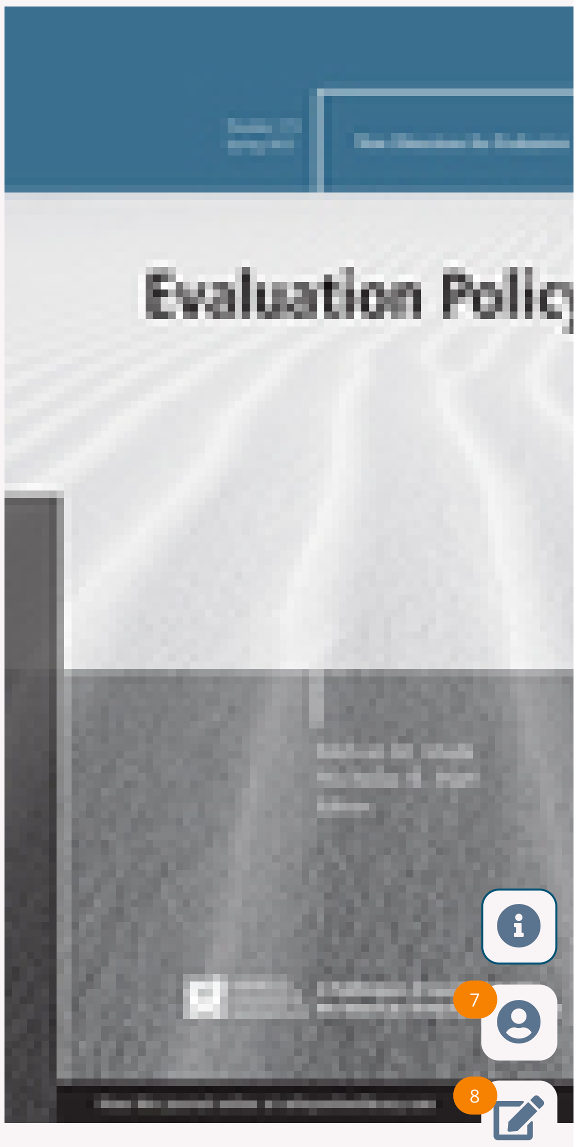
AQ4

- Inputs into strategic planning exercises (2 exercises)
- Inputs into multi-stage consultation process involving a wider range of civil society organizations (1 exercise)
- Content for papers published in academic journals (2 exercises)
- Findings for inclusion in evaluation reports of international aid agency policies and programs (2 exercises)
- Original research data for use in PhD studies (3, planned exercises)
- Refinement of a quality assurance process for humanitarian interventions (1exercise)
- Identification of alternative research uptake pathways that could be pursued in the next iteration of research funding (1 exercise)

The extent to which such expectations have been met is not always clear to the platform administrator. Some outcomes, such as published papers, become known when they are shared with the administrator. In other cases, it is simply not known. In many others, it is much more a matter of degree and inevitably involves subjective judgment. While there is no common evaluation approach that enables comparison of these exercises, there are some elements which are becoming more commonly practiced. One is the holding of post-exercise stakeholder workshops facilitated by the respective facilitators. These have been used to share evaluation findings and identify their implications (Futerman & Beard, 2023). The other is the publication of reports describing and assessing the completed exercises (Hutchings & Mugenzi, 2023).

Information on the process (e.g., activity) dimensions of ParEvo exercise is more immediately accessible by exercise facilitators. Data is continuously available to facilitators on the structure of participation—who built on who else's previous contribution. Figure 4 shows the distribution of participants' behavior on two axes: the extent to which they built on others' contributions and the extent to which others built on their contributions. To date this has been mainly used to track participants' tendency to build on their own contributions Such behavior could suggest a misunderstanding of the exercise if happening on a larger scale. But if happening on a smaller scale it may be just another useful source of diversity. Facilitators can also use a more gamified option by using a widget that shows participants a “leader board” of how they rank compared to others, in the extent to which others are building on their contributions. This has the potential to encourage the development of a more collectively owned set of storylines, though possibly at the cost of their overall diversity.

View PDF Redline PDF




7

8

1 Proof Initiated
7/17/2024

2 Corresponding Author
**Due date:** 7/19/2024
**Start date:** 7/17/2024
**End date:** 7/19/2024

3 Proof Collator
**Due date:** 7/21/2024
**Start date:** 7/19/2024

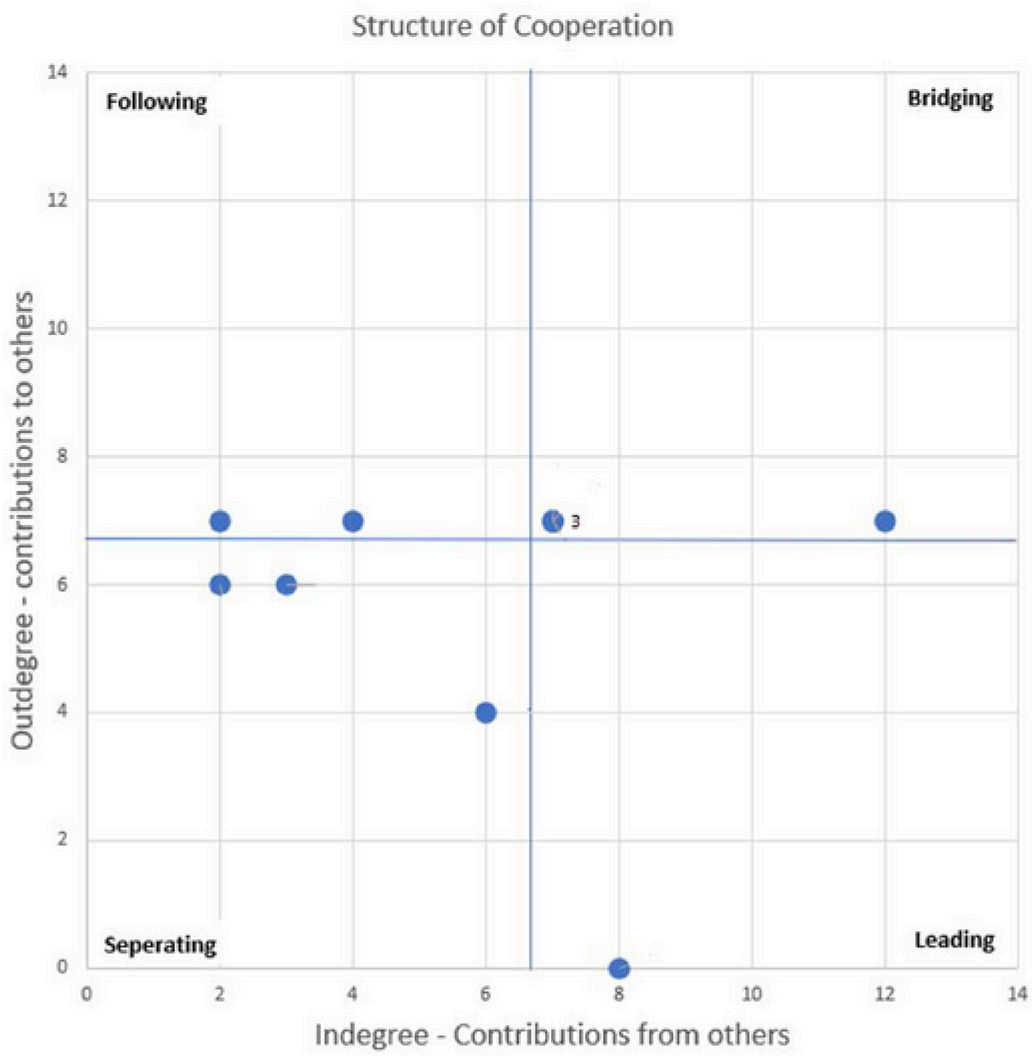



FIGURE 4.
The structure of cooperation, in the second CSER exercise. Gray dots = individual participants. Gray dividing lines = average possible for each participant. Four extremes of participant behavior are represented in each corner.[3]

## Platform level theories

The third type of theory of change is that concerning the functioning of the whole platform. In principle, ParEvo could be made available as a standalone downloadable software package, or simply as a printed set of workshop exercise instructions. However, the use of a shared online platform has multiple expected benefits. One is the quick and efficient updating of design features embedded in the software. Another, and more importantly, is the administrator's ability to access and compare data on multiple exercises. The objective here is to enable facilitators to carry out successful exercises based on all available experience to date, and to ensure a process of continuing improvement.

Much of the learning is incremental and opportunistic, identifying what works and what does not work in each new exercise. But there is also a larger ambition connected to the assumptions built into the design of the ParEvo process. The base assumption underlying the design of the ParEvo process is that the evolutionary algorithm (Dennett, 1996) is an effective learning process within complex environments, as evident from observation and experiments in biology and computer science (Hordijk, 2018). The evolutionary algorithm consists of three reiterated events: variation, selection, and reproduction. In ParEvo these are embodied as a structured social process, leading to a tree structure of narratives, within which some versions have not survived, and others have, and some of which have proliferated more than others. In both the biological world and in ParEvo, selection is an evaluative judgement, to do with a minimalist but important capacity to survive (Belew, 1991).

Within this design perspective, there are two associated assumptions about the role of diversity, that is, variation. Firstly, the diversity of participants is seen as an independent variable affecting the creativity of the process. This view of its causal role 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). Participants' diversity is measurable in their number, their types, and in the different ways they collaborate in the construction of different storylines, as shown in the Figure 4 example.

Diversity is also seen as a dependent variable that is descriptive of the range of possible futures that have been explored. The assumption here is that the 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. Diversity of storylines is measurable using diversity measures from ecology and more specifically the sociological uses of those ecological ideas: variety, balance, and disparity (Stirling, 2007). The diversity of storyline structures can also be measured using network visualization and associated measures, such as network density and closeness centrality (Borgatti et al., 2018).

Analyzing the consequences and causes of diversity within the ParEvo context is a long-term challenge. Because the administrator's intention has to encourage a wide range of parties to make use of ParEvo, each consecutive exercise is in effect 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, while others vary. The numbers of participants have been similar, as have the number of iterations completed. But the kind of futures being explored, and the associated guidance given to participants have been very different. They range from the alternative futures facing the Intersex Asia network to future global governance options for solar radiation management. In principle, the bivariate analysis of the relationship between the diversity of inputs (e.g., participant characteristics), activity (collaborative behavior), and outputs (storyline diversity) should not be technically demanding. But given the wide range of exercise topics and purposes a more configurational analysis that includes other exercise attributes will be more appropriate (Rihoux & Ragin, 2009). A necessary step in that direction that is currently underway is the continued accumulation of data from more cases and completed exercises.

## CONCLUSIONS

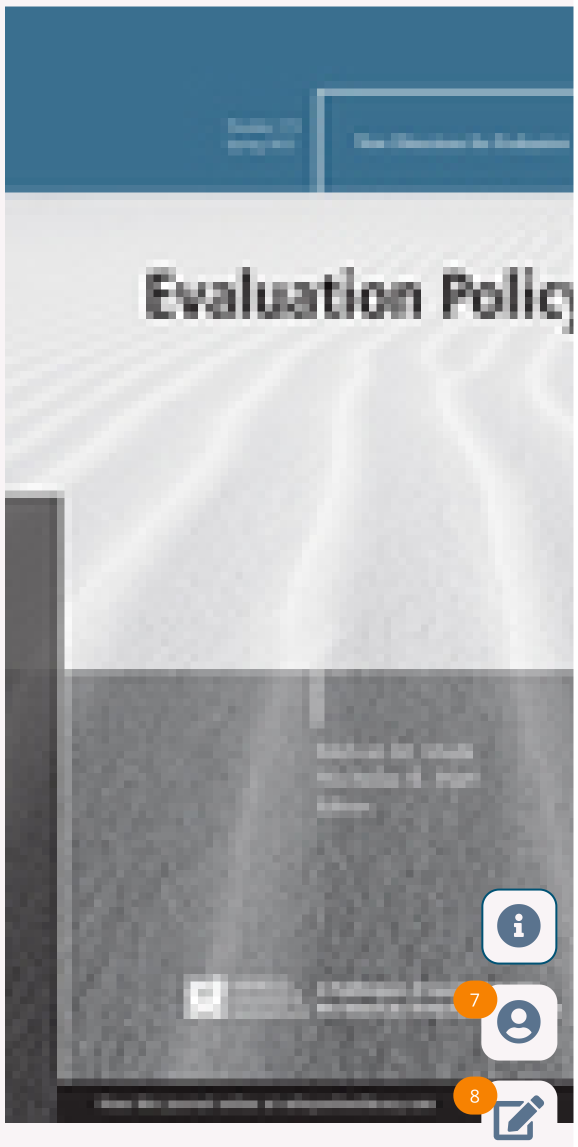
### What can be learned from this case example?

Single-case examples can show that particular events are possible in the real world and are not simply conjectures. In the case of ParEvo, what is evident is that a wealth of opportunities exists to build in evaluation activities within a foresight activity. These do not need to be confined to an explicit "Evaluation" event. Evaluative activities can be identified and shaped at multiple stages of a foresight exercise and thereafter, and they can involve different stakeholders in different roles.

Perhaps less evident is that evaluation can be conceived as a multi-level process. At the "ground level," the contents of the futures that are generated by any foresight exercise are all legitimately and usefully the focus of evaluative attention. At the next levels of scale, the aim and methods of those organizing those foresight exercises are also evaluable. Above that level are the ambitions of the platform administrator concerning its use and performance. In between, there is the possibility of analyses of multiple exercises, if they are being informed by a common overall strategy.

More generally, any foresight process that involves consideration of multiple alternative futures can be seen as involving generative and evaluative phases, similar to those seen within ParEvo (variation and selection). Alternative futures are articulated and these are selectively responded to in different ways. The evolutionary algorithm, as





implemented both in the biological and social worlds, is a search process that necessarily and implicitly involves some form of evaluation activity. The learning challenge is to make that evaluative activity transparent and accountable. This applies to both ParEvo exercises and many other forms of foresight exploring alternative futures.

For evaluators, it should be clear that foresight exercises such as ParEvo offer a much wider perspective on the future than is the case with most approaches taken to the development of theories of change for policies, programs, and projects. While the contents of likely and desirable futures will be relevant to the updating of existing theories of change, other futures outside these narrow bounds should be commanding evaluators' attention. Articulation of the significant differences between these alternative futures will help us identify “the box” we may find our thinking in at present, and thus the possibility of getting out of it.

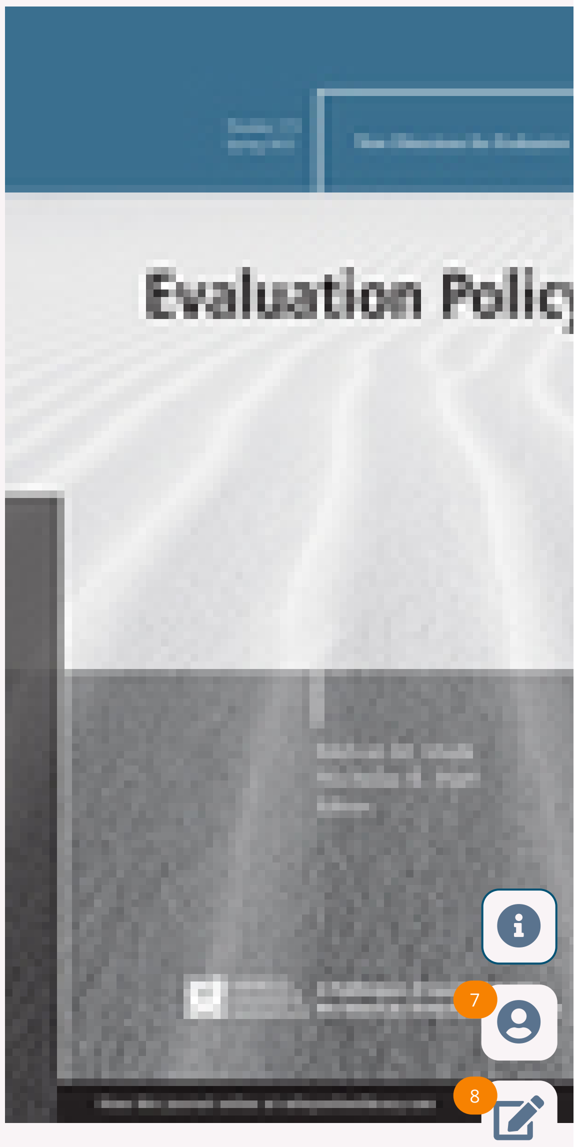
## LIMITATIONS AND WAYS FORWARD

The very fecundity of the ParEvo process can bring its own problems. ParEvo exercises to date have generated large bodies of text–around 15,000 words on average. Examples can be seen in the ICCS's publicly available completed exercise (Hutchings & Mugenzi, 2023). While some facilitators have extensive experience with content analysis methods, others have not and have had to rapidly learn by doing (Beard et al., 2023). In both cases, the time-consuming nature of many content analysis methods can limit the value that can be extracted from what are typically very rich “thick descriptions” of alternative futures (Geertz, 1972). However, the recent public availability of Large Language Model (LLM) forms of Artificial Intelligence, now offers a much quicker and more customizable way of analyzing and comparing the contents of ParEvo storylines (Head et al., 2023). Promising results have already been made with the generation of headline summaries of participants' contributions and the identification of significant differences between the contents of different groups of storylines, e.g., surviving versus extinct storylines.

As may already be evident, embedded in any foresight exercise are a multitude of judgments by different stakeholders about what to do next. For example, in each iteration of a ParEvo exercise, participants choose which storyline they want to build on. They may do so for a range of reasons, from randomly to systematically applied. These choices are the causal mechanisms driving the evolution of the storylines. However, in the ParEvo exercises to date, they have received relatively little attention. Yet they relate to a significant debate in evolutionary theory about the role of random forces (known as genetic drift) versus consistent environmental pressures of one kind or another (Le Page et al., 2020). Knowledge of any conscious influences at work could shed more detailed light on how participants are thinking about alternative futures, and give facilitators the option of providing informed guidance on the selection of storylines.

The management structure of the ParEvo.org platform also brings limitations as well as opportunities. The ability of different self-selected individuals and organizations to independently run their own exercises on the platform means a wide range of exercise purposes and settings have already been explored, well beyond the domain knowledge and social networks of the administrator. However, the facilitators' very independence means that the comparability of the processes and outcomes of those exercises is partial. Facilitators interest, knowledge, and resources available to invest in exercise evaluations also vary. However, the background technical support offered to all facilitators by the administrator means that sharing of knowledge across exercises has





been possible on a case-by-case basis. As the number of completed ParEvo exercises grows so will the potential for more systematic testing of hypotheses and extraction of lessons through cross-case analyses.

The experience of enabling the ParEvo process to be run via an online platform may also prompt more useful experiments with this approach to making particular evaluation methods both available and comparable. Two other similar platform-based initiatives are already underway, led by the author. EvalC3, an Excel app developed for small-scale prediction modeling and case-based inquiry is being developed into a publicly available web-based application by Sheffield Hallam University. Another web application enabling structured ethnographic inquiries will also soon become available to evaluators on another dedicated platform.

## CONFLICT OF INTEREST STATEMENT

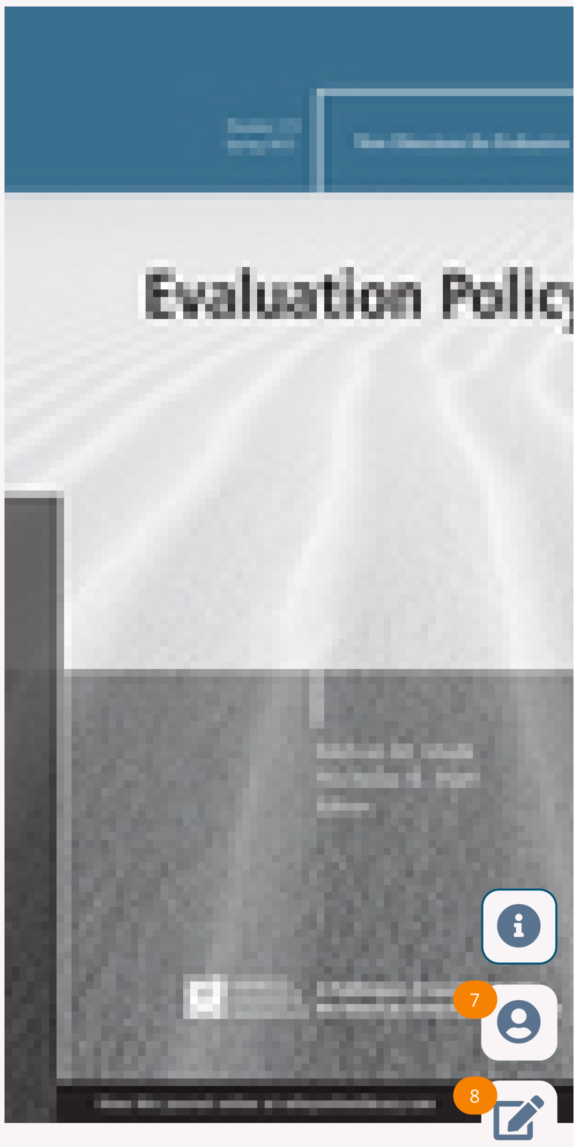
The author declares no conflicts of interest.

### ENDNOTES

1 Facilitators are self-selected people who want to run a ParEvo exercise, and who register to do so on www.paraevo.org. They are responsible for the design of their exercise, including selection of participants. They are provided with background technical support by the platform administrator (the author), before and during their exercise.

2 The participant survey also asks participants about the events they think are most surprisingly missing from and included in the storylines

3 (a) Leading – building on one's own contributions and others building on these, (b) Isolating – building on one's own contributions and others not building on these, (c) Following – building on others contributions, and others building on these, and (d) Bridging – building on others contributions and other building on one's own





## REFERENCES

Beard, S. J., Cooke, N., Dryhurst, S., Cassidy, M., Gibbins, G., Gilgallon, G., Holt, B., Josefina, I., Kemp, L., Tang, A., Julius, W., Ingram, P., Sundaram, L., & Davies, R. (2023). *Exploring futures for the science of global risk* (SSRN Scholarly Paper 4405991). Retrieved from https://papers.ssrn.com/abstract=4405991

Belew, R. K. (1991). Artificial life: A constructive lower bound for artificial intelligence. *IEEE Expert*, *6*(1), 8–15.

Bishop, P., Hines, A., & Collins, T. (2007). The current state of scenario development: An overview of techniques. *Foresight*, *9*(1), 5–25. Retrieved from https://www.researchgate.net/publication/228623754_The_current_state_of_scenario_development_An_overview_of_techniques

AQ5 Blundo-Canto, G., Ferré, M., Tibaduiza-Castaneda, L., Andrieu, N., Fréguin-Gresh, S., Rodriguez-Borray, G., Santacruz-Castro, A., & Mercandalli, S. (2024). Adapting and combining foresight and ex ante impact pathway evaluation for place-based research planning with stakeholders. *New Directions in Evaluation*, *2024*(181), 00–00.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks*. Sage.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, *67*(6), 380–400. https://doi.org/10.1037/h0040373

Campbell, C. M., Izquierdo, E. J., & Goldstone, R. L. (2022). Partial copying and the role of diversity in social learning performance. *Collective Intelligence*, *1*(1), 26339137221081849. https://doi.org/10.1177/26339137221081849

Carden, F. (2023). Back to the future: Are we trapped in our past? *Evaluation and Program Planning*, *97*, 102264. https://doi.org/10.1016/j.evalprogplan.2023.102264

Clark, H. (2022). *TOCO 3.0*. ActKnowledge theory to results. Retrieved from https://www.actknowledge.org/toco-software/

Conrad, P., Sarva, S., & Smith, R. (2023). *Choosing your tomorrows: Using foresight and anticipatory governance to explore multiple futures in support of risk-informed development*. UNDP. Retrieved from https://www.undp.org/publications/dfs-choosing-your-tomorrows-using-foresight-and-anticipatory-governance-explore-multiple-futures-support-risk-informed-development

AQ6 Dart, J., & Gates, E. (2024). Incorporating futures thinking into theory of change: Case and lessons learned from a social enterprise intermediary in Australia. *New Directions in Evaluation*, *2024*(181), 00–00.

Davies, R. (1998). *Order and diversity: Representing and assisting organisational learning in non-government aid organisations* [Doctoral dissertation, University of Swansea, Wales]. Retrieved from https://www.researchgate.net/publication/333894025_Order_and_Diversity_Representing_and_Assisting_Organisational_Learning_in_Non-Government_Aid_Organisations

Davies, R. (2019). *ParEvo: A web-assisted process enabling the participatory exploration of alternative futures*. Retrieved from https://mscinnovations.wordpress.com/

Davies, R. (2020, January 15). List of ParEvo exercises. *ParEvo*. Retrieved from https://mscinnovations.wordpress.com/list-of-parevo-exercises/

Davies, R. (2022, March 17). Evaluation questions. *ParEvo*. Retrieved from https://mscinnovations.wordpress.com/evaluation-questions/

Davies, R. (2023, August 31). Evaluating thematic coding and text summarisation work done by artificial intelligence (LLM). *Rick on the Road*. Retrieved from https://mandenews.blogspot.com/2023/08/

Davies, R., Hobson, T., Mani, L., & Beard, S. (2023). ParEvo: A methodology for the exploration and evaluation of alternative futures. *Evaluation*, *30*(1), 27–49. 13563890231188743. https://doi.org/10.1177/13563890231188743

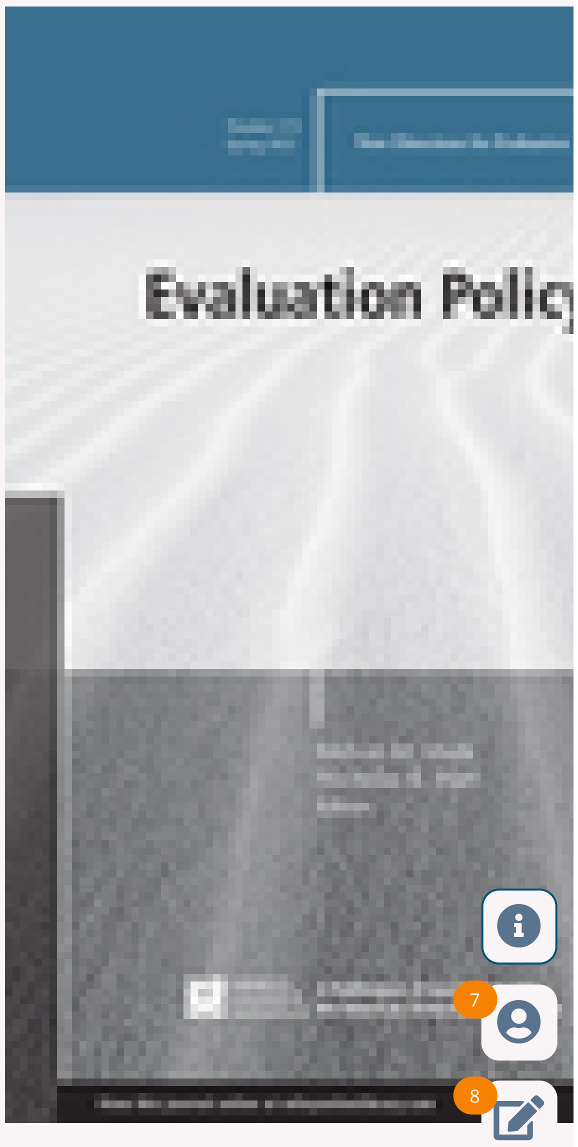
Dennett, D. C. (1996). *Darwin's dangerous idea: Evolution and the meanings of life*. Penguin UK.

DFID. (2011). *DFID's Approach to value for money (VfM)*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/49551/DFID-approach-value-money.pdf

Futerman, G., & Beard, S. J. (2023). *Managing the contribution of SRM and climate change to GCR* (p. 29). Retrieved from https://www.cser.ac.uk/resources/workshop-report-managing-contribution-srm-and-climate-change-gcr/

Geertz, C. (1972). Deep play: Notes on the balinese cockfight. *Daedalus*, *101*(1), 1–37. Retrieved from

View PDF Redline PDF



7
8

1 Proof Initiated
7/17/2024

2 Corresponding Author
**Due date:** 7/19/2024
**Start date:** 7/17/2024
**End date:** 7/19/2024

3 Proof Collator
**Due date:** 7/21/2024
**Start date:** 7/19/2024

https://www.jstor.org/stable/20024056

AQ7

Halonen, M., Hyytinen, K., & Kukri, S. (2024). Evaluation of futures frequency method: Approach and main results. *New Directions in Evaluation*, *2024*(181), 00–00.

Hamel, S., Ims, R., & Yoccoz, N. (2021). Challenges and opportunities when implementing strategic foresight: Lessons learned when engaging stakeholders in climate-ecological research. *Climate Research*, SUSTAIN, *86*, 29–35. https://doi.org/10.3354/cr01653

Head, C. B., Jasper, P., McConnachie, M., Raftree, L., & Higdon, G. (2023). Large language model applications for evaluation: Opportunities and ethical implications. *New Directions for Evaluation*, *2023*(178–179), 33–46. https://doi.org/10.1002/ev.20556

Hejazi, A. (2023). *A step-by-step approach toward becoming a professional futurist*. APF. Retrieved from https://www.apf.org/post/a-step-by-step-approach-toward-becoming-a-professional-futurist

Hordijk, W. (2018). *Evolution as a problem solver in computer science*. Retrieved from https://www.prosocial.world/posts/evolution-as-a-problem-solver-in-computer-science

Hutchings, H., & Mugenzi, P. (2023). *A history of civic space, 2024–2034*. Retrieved from https://solidarityaction.network/wp-content/uploads/a-history-of-civic-space-2024-2034.pdf

Le Page, M. L., Barras, C., Webb, R., Douglas, K., & Arnold, C. (2020). Evolution is evolving: 13 ways we must rethink the theory of nature. New Scientist. Retrieved 25 September 2020, from https://www.newscientist.com/article/mg24733010-800-evolution-is-evolving-13-ways-we-must-rethink-the-theory-of-nature/

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87. Retrieved from https://www.jstor.org/stable/2634940

McAfee, A., & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future*. W.W. Norton.

Poli, R. (2018). A note on the classification of future-related methods. *European Journal of Futures Research*, *6*(1), 15. https://doi.org/10.1186/s40309-018-0145-9

Popper, R. (2010, April 6). 33 Foresight methods. *Rafael Popper's blog*. Retrieved from https://rafaelpopper.wordpress.com/foresight-methods/

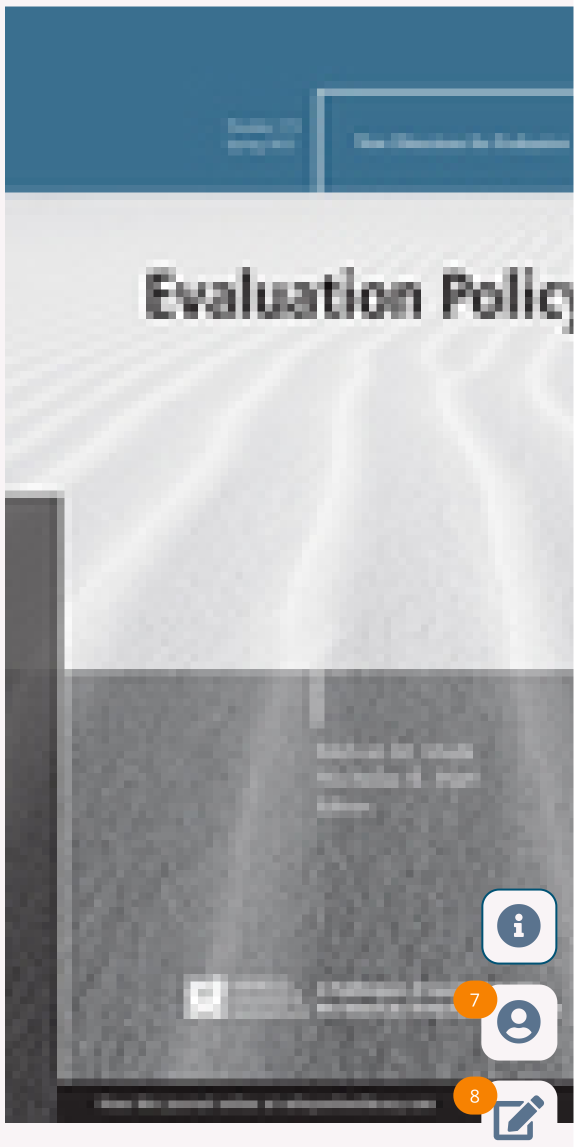
Raford, N. (2010). *Harnessing collective intelligence for crowdsourced scenario planning*. Retrieved from http://noahraford.com/?p=462

Rihoux, B., & Ragin, C. C. (Eds.). (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Sage.

Rhydderch, A. (2017). *Scenario building: The 2x2 matrix technique*. Retrieved from https://www.researchgate.net/profile/Alun-Rhydderch/publication/331564544_Scenario_Building_The_2x2_Matrix_Technique/links/5c80fc83458515831f8b5508/Scenario-Building-The-2x2-Matrix-Technique.pdf

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, *4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213


View PDF Redline PDF



7

8

1 Proof Initiated
7/17/2024

2 Corresponding Author
**Due date:** 7/19/2024
**Start date:** 7/17/2024
**End date:** 7/19/2024

3 Proof Collator
**Due date:** 7/21/2024
**Start date:** 7/19/2024

Waverley Consultants. (2017). *The futures toolkit: Tools for futures thinking and foresight across UK government*. Government Office for Science. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/674209/futures-toolkit-edition-1.pdf

Wikipedia. (2024). Backcasting. In *Wikipedia*. Retrieved May 12, 2024 from https://en.wikipedia.org/w/index.php?title=Backcasting&oldid=879511783

Wilden, R., Hohberger, J., Devinney, T. M., & Lavie, D. (2018). Revisiting James March (1991): Whither exploration and exploitation? *Strategic Organization*, *16*(3), 352–369. https://doi.org/10.1177/1476127018765031

Woo, A. (2017). *To think outside the box, You first need to find the box*. LinkedIn. Retrieved from https://www.linkedin.com/pulse/think-outside-box-you-first-need-find-annie-woo/

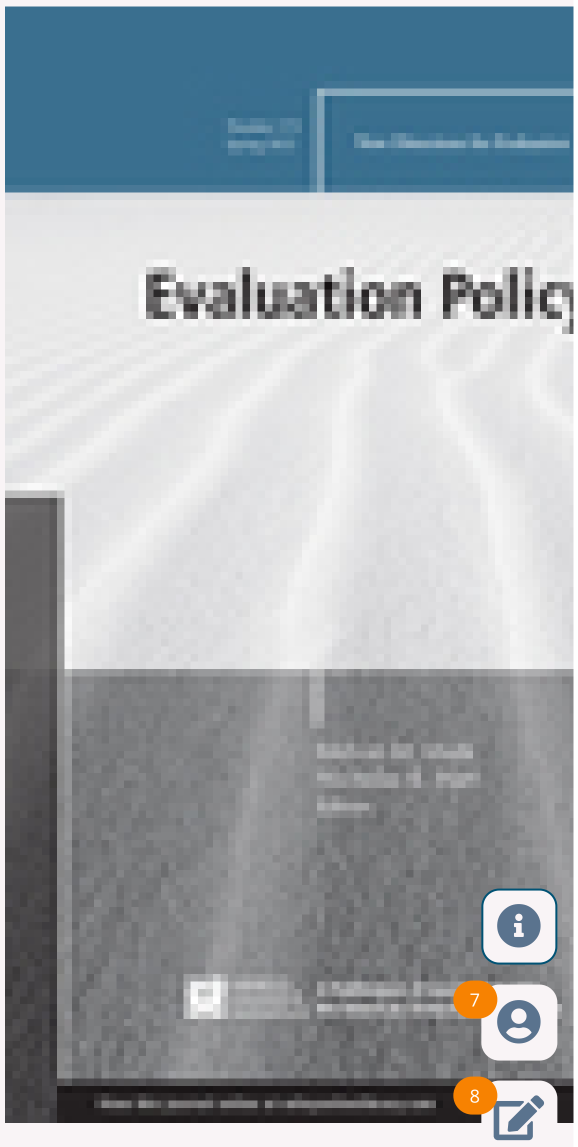
## AUTHOR BIOGRAPHY

**Rick Davies**, PhD, is an independent Evaluation Consultant based in Cambridge, United Kingdom.

**How to cite this article:** Davies, R. J. (2024). Integrating foresight and evaluation: The example of ParEvo. *New Directions for Evaluation*, 1–14. https://doi.org/10.1002/ev.20606

-----------------------End of Document-----------------------

View PDF Redline PDF



1 Proof Initiated
7/17/2024

2 Corresponding Author
**Due date:** 7/19/2024
**Start date:** 7/17/2024
**End date:** 7/19/2024

3 Proof Collator
**Due date:** 7/21/2024
**Start date:** 7/19/2024