# If this, then that, then what? A generative process for overcoming implicit bias in the initial phases of participatory modelling

**C.A. Browne and E. Nabavi**

*Responsible Innovation Lab, The Australian National University, Canberra*
*Email: Chris.Browne@anu.edu.au*

**Abstract:** Model elicitation is the process by which the knowledge provided by participants is translated into a model that is generated by skilled modelers for gaining insight into a given problem. In this paper, we explore participatory modelling methodology championed within the system dynamics community. One of the challenges in working in participatory approaches for model elicitation is the limited time available with stakeholders to work through the complex causal relationships that exist in the often-diverse mental models within a group of participants. Specifically, these challenges can lead to significant bias in the modelling process, often leading to suboptimal results.

To help overcome these biases, we propose a simple generative methodology to help participants provide a base to explore many alternative hypotheses. The result is not a single model that a participant or learner is anchored to, but rather a collection of causal structures which can help navigate in the initial stages of an investigation the complex inter-relationship of dynamic variables. The process, which we have called the 'If this, then that, then what?' technique, encourages the systematic consideration to overcome the bounded rationality that exists in mental models within a time-sensitive environment. The structure is simple, iterative, generative, and designed overcome the bounded rationality we know exist within mental models.

In an experiment to test the approach, 30 participants were guided through a 45-minute workshop that encouraged structured brainstorming to consider the likely positive, neutral, and negative effects of a simple intervention. In this activity, over 131 causal stories were created which could be mapped to a problem space involving 34 combinations of overlapping thematic areas. This significantly expands the opportunities for model exploration by exploring perceived dynamic behaviour prior to model creation. Further, this approach has the potential to inform how causal feedback models are created, especially with novice modellers.

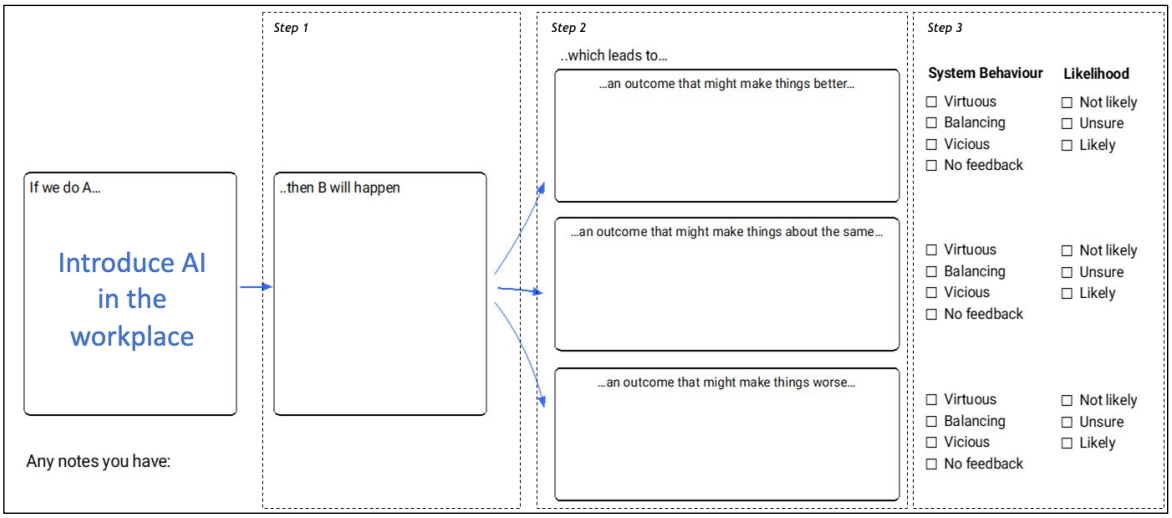


**Figure 1.** Template for structured causal thinking in the 'If this, then that, then what' technique. Note: 'steps' indicated by the dotted boxes represent the steps shown in Table 2 and were not included on the worksheet

***Keywords:*** *Systems thinking, participatory modelling, causal loop diagrams, cognitive bias*

## 1. INTRODUCTION

The practice of model elicitation involves the translation of domain, problem and non-scientific participant knowledge into the representation of a model generated by expert modelers to gain insight on a given problem. In this case, we explore the methodological processes championed within the system dynamics community. Group model building and participatory modelling approaches typically bring together domain experts with limited knowledge of modelling together with modelling experts to explore and gain insights about the complex dynamics of a given situation (Andersen and Richardson 1997; Andersen et al. 1997; Hovmand 2014; Hovmand et al. 2012; Vennix 1996). As such, participatory modelling approaches are an important technique for working with domain experts, promoting inclusivity in collaborative processes, and an opportunity to teach the art and science of modeling.

One of the challenges in working in participatory approaches for model elicitation is the limited time available with stakeholders—often driven by external factors such as availability of domain experts—to work through the complex causal relationships that exist in the often-diverse mental models within a group of participants. Mental models represent a subset of the real world, are built from experience, and are incomplete and therefore difficult to simulate (Maani and Cavana 2007; Meadows 2008; Sterman 2000). Because of this, it is often challenging for participants and facilitators to share their mental models about a given situation in an effective and complete way.

Due to the limited time that participatory modelers have with participants, it is commonplace for participatory modelers to use devices such as scripts, worksheets, and templates to quickly elicit the information required to construct models that comply with agreed conventions (Hovmand et al. 2012; Scott et al. 2013; Scriptapedia Wikibooks contributors n.d.). These devices often enable participants to develop a shared understanding of a dynamical situation through a series of explanatory, divergent, convergent and evaluation tasks.

Consider a participatory modelling process that is eliciting the mental models of participants around an agreed problem. A common early step is to determine the first simple causal inference: *when A happens, B will happen.* For example, *if funding increases, we can expect better service.* This initial causal relationship frames the rest of the modelling process. Consider the same relationship framed from a negative perspective: *if funding decreases, we can expect worse service*. These frames are useful to capture the perspectives of the participants, but the expert modeler would recognise that the relational structure of these two statements is identical: *the level of funding influences the quality of service*, with a positive (same) polarity. This example describes but one of countless concerns in translating mental models into causal relationships.

Time factors, the social nature of participatory modelling, the very problem framing at hand, the inherent issues around heuristics of decision-making (Atkinson et al. 2015), and implicit bias of modelers themselves (Größler 2004; Sterman 2000) can introduce significant issues concerning forms of bias when these are brought into a participatory modelling process (Hoch et al. 2015). Such bias can manifest as 'social biases'—such as groupthink, arriving at a false consensus and bandwagon effects—'time shortcuts'—such as availability heuristics (Kahneman 2011)—and 'cognitive limitations'—such as bounded rationality, anchoring effects and exploring unintended consequences. These forms of bias are often not addressed in the early stages of model building, and then are embedded as core truths as the modeler builds out the complexity of the model. For the expert modeler facilitating the process, this can lead to the wasted time and effort of *building an insightful model of the wrong problem*. This highlights the need for approaches to embed responsible modelling practices into participatory approaches.

In the following sections we describe a methodology that can be used to overcome many of these biasing effects in model conceptualisation, and the results from an initial experiment.

## 2. APPROACH

To help overcome some of the implicit biases that can occur in participatory modelling, we have developed a simple generative methodology to help participants provide a base to explore many alternative hypotheses. The result is not a single model that a participant or learner is anchored to, but rather a collection of causal structures which can help navigate the complex inter-relationship of dynamic variables. The process, which we have called the 'If this, then that, then what?' technique, encourages the systematic consideration of alternative hypotheses to help overcome the bounded rationality that exists in mental models within a time-sensitive environment. Further, it also stimulates 'reflexivity' among participants and modelers and contributes to responsible modelling practices (Nabavi 2022; Stilgoe et al. 2013).

The structure is simple, iterative, generative, and designed to overcome the bounded rationality we know exists within mental models. The main activity can be characterised in four steps, shown in Table 1.

**Table 1.** Four steps characterised in the 'If this, then that, then what?' activity

| Step and Prompt | Description |
|---|---|
| **Step 1**<br>**If this, then that...**<br><br>*Consider a causal link: if A happens, then B will happen.* | This step is undertaken individually. It is the simplest unit of causal hypothesis that can be used to construct a causal relationship. It is typically this basic causal reasoning that a policy is built on: when 'A' happens, then 'B' will happen.<br>Our experience is that the most value is derived through an agreed initial link informed by a validated, reliable source to provide a common starting point in the causal chains that are developed in subsequent steps. |
| **Step 2:**<br>**…then what?**<br><br>*Consider alternative futures: if A happens, then B will happen, which leads to...* | This step is also undertaken individually. Participants consider alternative futures that might arise from the initial causal link. These build alternative causal chains of reasoning. The provided prompts are to consider:<br>• an outcome that might make things better<br>• an outcome that might make things about the same<br>• an outcome that might make things worse<br>The three outcomes are proxies for the system behaviour that might arise from the developing structures. It is likely that structure of a model that pushes the system behaviour from its current state (better or worse) may show reinforcing feedback, whereas about the same may show balancing feedback. Although in causal model structures, the better or worse behaviour derives from the same structure, we find it useful to consider these behaviours separately in this phase.<br>This positioning of different possible outcomes from the same initial causal link also helps to overcome false consensus and anchoring bias, as the activity actively encourages participants to think about alternative futures and explore the unintended consequences of their initial position. |
| **Step 3:**<br>**Look for feedback.**<br>*Draw circular connections* | In this step, undertaken individually or in pairs, participants develop the alternative futures from Step 2 into feedback loops where possible, and to identify the behaviour as Balancing or Reinforcing. We also ask participants to identify whether the labelling of Reinforcing loops encourages virtuous (i.e., better and better) or vicious (i.e., worse and worse) behaviour, and to consider how they could rephrase their variables to accurately represent the causal loop for both. These loops become the basis for a broader discussion. |
| **Step 4:**<br>**Discussion and sharing.**<br>*Share understanding across groups* | The final step is to use the alternative models to foster discussions within groups. If the participatory modelling process allows, there may be opportunities to discuss and share the models between different groups, or to explore combinations of feedback structures within the groups. This becomes the base point for further investigation, such as exploring the validity of causal links through collection of data, agreeing, or disagreeing on conflicting positions, or a useful record of initial thinking for a modeler to use and build out models further. |

The steps in the activity could be undertaken individually, as part of a group activity, or in combination. This divergent approach results in 'many models' rather than one model, all exploring perspectives and unintended consequences of the same fundamental causal link. The concept of developing many causal hypotheses rather than one is one way to escape the bounded rationality inherent in mental models. Additionally, it can allow participants to celebrate the diversity inherent in different perspectives, and to develop a more complete understanding of the problem space.

## 3. METHODOLOGY

To explore the applicability of the process described in §2, an initial experiment was conducted through a participatory modelling project with 30 postgraduate students enrolled in a professional practice course in the disciplines of engineering and computing during a regular class. Participants were not required to complete the activity as part of class, were not graded on the activity, and were otherwise not incentivised for participation. For the purposes of the experiment, all participants were asked to consider the example of introducing artificial intelligence (AI) into the workforce in a structure shown in Table 2. However, the goal of the workshop was to have students learn a simple systems-thinking process to help them consider the unintended consequences of a concurrent but unrelated project within the course.

**Table 2.** Overview of timing used in the experiment (total time approximately 45 min)

| **Time** | **Description** | **Actor/s** |
|---|---|---|
| 10 min | Introduction to systems thinking and overview of experiment | Researchers |
| 5 min | Introduction to problem statement and sharing of news video | Researchers |
| 5 min x 3 | Brief explanation of step; time for participants to complete worksheet step, for each of the 3 steps | Individuals |
| 10 min | General discussion between groups about different answers and individual reflection | Small groups |
| 5 min | Concluding remarks | Researchers |

Participants completed an individual worksheet shown in Figure 1 throughout the experiment, which were collected and used to generate the data for analysis. Due to issues of the variety of textual descriptions provided in the free-text responses, these were categorised into the substantive themes for the analysis in §4.

Responses were transcribed from steps 1-3 in the original worksheets into a spreadsheet for analysis. All responses were in relation to the prompt of 'if we introduce AI in the workplace'. Free-text responses were coded manually using an inductive process with flat coding frame based on the responses within the data set. Sentiment was removed; for example, a response describing 'losing jobs' would be categorised into 'employment'. Where participants provided multiple statements that spanned codes, these statements were split into multiple entries; for example, the response '*Increase productivity but may make people unemployed*' was coded as both 'employment' and 'productivity'. The step 1 responses were separated into five categories, and the step 2 responses were separated into eight categories. Category descriptions and examples are shown in Table 3.

**Table 3.** Description of coded categories in free-text responses

| Step | Code | Description | Example response/s |
|---|---|---|---|
| 1 | automation | concerning activities that result in shifting tasks from humans to AI | "AI will take over certain tasks of a process"<br>"A lot of tasks will be automated […] translation, editing" |
| 1,2 | efficiency | concerning the speed or time to undertake a task | "Make work more efficient"<br>"Increased performance efficiency" |
| 1,2 | employment | concerning jobs or replacing humans | "Some people will lose their jobs (replaced by AI)"<br>"May cause some sort of job cut & unemployment" |
| 1,2 | productivity | concerning volume or processes of output | "Improve the productivity in general"<br>"[...]increasing the productivity of resources." |
| 1,2 | quality | concerning the quality of output or performance | "The quality of work will be the same comparing to human"<br>"Customers are not satisfied by AI […] services." |
| 2 | errors | concerning creation or identification of errors | "Mistakes made by AI are sometimes are inevitable"<br>"AI […] process produce a mistake in its working algorithm." |
| 2 | finance | concerning costs, payments, or profits | "Company shifts cost from salary to AI services charges."<br>"Increase profit" |
| 2 | outputs | concerning the output or process itself | "AI doesn't improve the product itself."<br>"More standard work procedures" |
| 2 | scope of work | concerning the changing nature of employment or task allocation | "Some skilled jobs still need to be done by human"<br>"Open opportunities for work/life balance" |

## 4. RESULTS

Thirty participants recorded responses during the workshop, 9 with multiple responses. The 39 responses described in step 1 resulted in 131 causal chains, which could be clustered into 34 categories between step 1 and step 2 responses. In some cases, no response was captured, which has been reported as 'nil response'. A summary of descriptive results by workshop step is shown in Table 4, with additional relationships between responses shown in Figure 2.

**Table 4.** Summary of results by workshop step

| Step | Response | Count | (%) |
|---|---|---|---|
| 1<br>'then this'<br>free text<br>number of response = 39 | employment | 13 | (33) |
| | automation | 11 | (28) |
| | efficiency | 10 | (26) |
| | productivity | 3 | (8) |
| | quality | 2 | (5) |
| 2<br>'then what'<br>free text<br>number of causal chains = 131<br>nil response = 2 | employment | 34 | (26) |
| | scope of work | 25 | (19) |
| | efficiency | 22 | (17) |
| | outputs | 12 | (9) |
| | quality | 12 | (9) |
| | errors | 11 | (8) |
| | productivity | 8 | (6) |
| | finance | 5 | (4) |

| Step | Response | Count | (%) |
|---|---|---|---|
| 2<br>'then what'<br>direction<br>n = 131 | better | 46 | (35) |
| | same | 43 | (33) |
| | worse | 42 | (32) |
| 3<br>system behaviour<br>n = 131 | virtuous | 45 | (34) |
| | balancing | 30 | (23) |
| | vicious | 34 | (26) |
| | no feedback | 9 | (7) |
| | nil response | 13 | (10) |
| 3<br>likelihood | likely | 81 | (62) |
| | unsure | 25 | (19) |
| | not likely | 6 | (5) |
| | nil response | 19 | (15) |

The response categories provide a broad view of the problem through the collective generation of ideas. Figure 2 shows the relationships between responses between steps 1-3 in the form of a Sankey diagram, which visually relates the connections in the responses. In this form, the relative widths of connecting lines represents the number of responses in each category, providing a sense of the issues and factors that participants were concerned about at the time of the workshop.

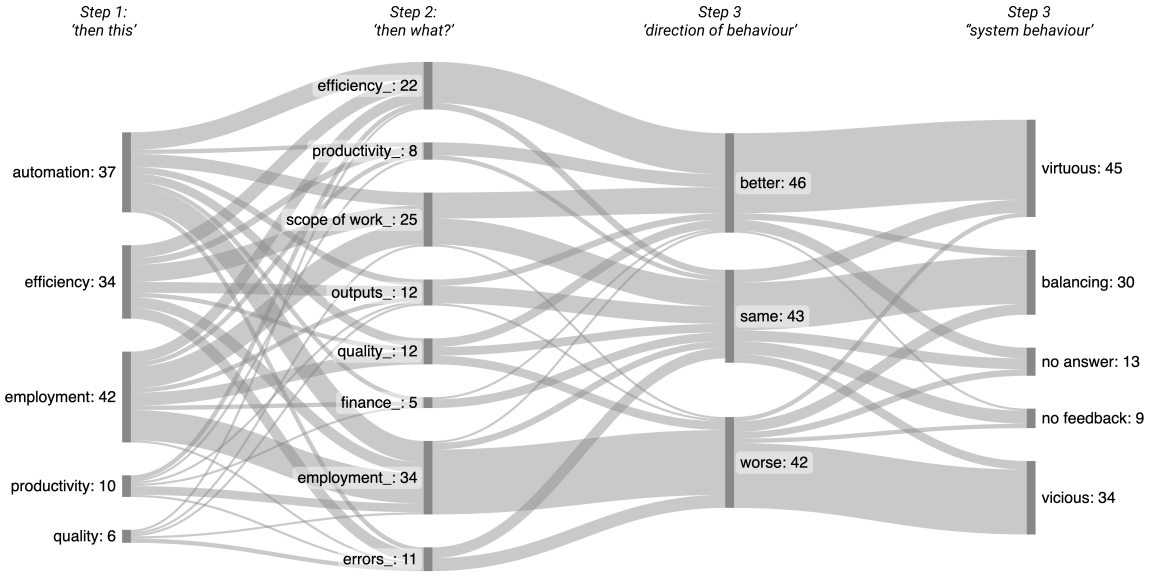


**Figure 2.** Visualisation of responses by workshop step

## 5. DISCUSSION

The relationships shown in Figure 2 provide a rich overview of themes arising from the introduction of AI into the workplace. The benefit of undertaking a generative process such as this is that we have at our disposal 131 causal chains generated from 30 participants prior to the first causal link being created in any shared model, immediately prompting broader thinking to overcome many forms of bias in the model building process. However, care is required to ensure that this process does not introduce new forms of bias: for example, simply working through the dominant links may amplify the groupthink that may have emerged from within the modelling group, or the exploration of possibilities early in the process might limit thinking later in the process about more complex system behaviour, such as oscillation, tipping points and phase changes.

However, within the scope of this work, there are some noticeable trends in Figure 2 that represent the collective mental models of the participants that could be used to prompt or prime discussion prior to commencing a formal model. To illustrate, we work from right to left in the figure:

*Q: What trends are there between the direction of behaviour and the system behaviour?*

Example: The 'better' responses tend to exhibit 'virtuous' behaviour; the 'worse' responses tend to exhibit 'vicious' behaviour; 'same' responses tend to exhibit 'balancing' behaviour. These feedback dynamics may indeed be working against each other or balancing efforts. Delays between these feedback loops may lead to complex system behaviour.

*Q: What trends are there between our 'then whats' and the direction of behaviour.*

Example: Looking at the dominant trends, we can see dominant relationships such as: 'employment' to 'worse', 'efficiency' to 'better', 'scope of work' to 'better' and 'same'. What are the stories with these frames, and are there frames that have not been considered or represented? Such as how the introduction of AI may lead to more employment (such as through new industries).

*Q: What trends are there between 'then this' and 'then whats'*

Example: Take any of the 'then this' variables and follow them through looking for plurality of views. See, for example, that 'automation' influences 'efficiency' in a positive sense, 'employment' in a negative sense, and 'scope of work' in both a positive and neutral sense. Explore the stories that arise and start to build up a narrative from the individual responses.

With the combinatory complexity of the coded categories, the direction of behaviour and the system behaviour, it could be tempting to jump right in and build a model that brings together these factors. However, in the context of a participatory modelling workshop, we suggest at this point that there is value in exploring many small models, and to use the collective insights from this early phase to inform the model building process. This process, visually shown in Figure 3, could look like:

***Choose a 'then this' to focus on, or divide them between groups, and express it as a causal link.***

For example, consider the link between introduction of a) AI and b) automation.

***Explore 'then what' factors to generate separate feedback structures in 1-loop systems.***

For example, explore the reinforcing (R) and balancing (B) feedback that may arise between a) AI, b) automation, and c) such as employment, efficiency, or scope of work.

***Explore further 'then what' stories to generate rich alternative stories in 2-loop systems.***

For example, taking a), b), and c) above, what further variables (i.e., d, e, f) might interact in combinations of reinforcing and balancing loops (i.e., RR, RB, BR, BB).

This simple scaffold cascades from one causal link to two 1-loop systems and four 2-loop systems, which can then be explored for connections and relationships between models and between participating groups.

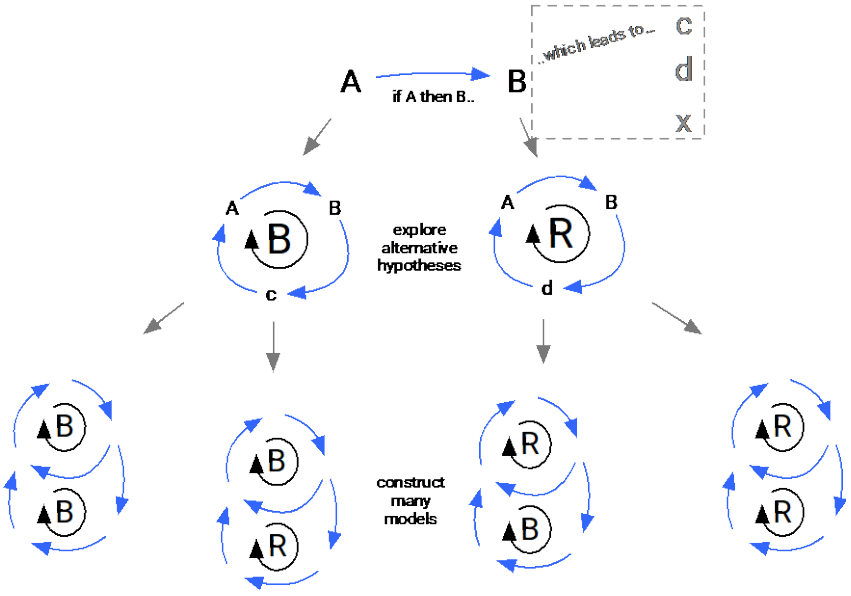


**Figure 3.** Cascading many models from the initial causal link

## 6. CONCLUSION

In this paper we have described a process developed to help facilitate the construction of models in a participatory setting and overcome some forms of bias that are frequently observed in these settings. The 'If this, then that, then what?' structure is a simple scaffold that encourages participants to explore multiple hypotheses in the problem space. An experiment was undertaken with novice modellers to explore the extent to which multiple causal chains could be created in a short time. In the space of approximately 45 minutes, one causal link led 30 participants to generate 131 causal hypotheses, which were categorised into 34 groups. This demonstrates the potential of this structured approach to broaden the shared mental models of participants prior to engaging in model building activities, such as in participatory modelling processes. A structured approach to exploring the broad problem space was proposed which builds out models one feedback loop at a time to encourage the participatory modellers to explore 'many models'. Further work is required to explore the effectiveness of this technique beyond the scope of the experiment set at the initial phase.

We have found that this process is a simple scaffold for non-expert modellers, such as those in participatory settings or students starting to model feedback systems, to develop an understanding of fundamental feedback structures, promote diversity of thought within a participatory modelling process, and uncovers potential unintended consequences of an initial causal hypothesis.

## REFERENCES

Andersen, D., Richardson, G.P., 1997. Scripts for group model building. System Dynamics Review 13, 107–129. https://doi.org/0883-7066/97/020107-23

Andersen, D.F., Richardson, G.P., Vennix, J.A., 1997. Group model building: adding more science to the craft. System Dynamics Review 13, 187–201.

Atkinson, J.-A., Wells, R., Page, A., Dominello, A., Haines, M., Wilson, A., 2015. Applications of system dynamics modelling to support health policy. Public Health Research & Practice 25. http://dx.doi.org/10.17061/phrp2531531

Größler, A., 2004. A content and process view on bounded rationality in system dynamics. Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research 21, 319–330.

Hoch, C., Zellner, M., Milz, D., Radinsky, J., Lyons, L., 2015. Seeing is not believing: cognitive bias and modelling in collaborative planning. Planning Theory & Practice 16, 319–335. https://doi.org/10.1080/14649357.2015.1045015

Hovmand, P.S., 2014. Group Model Building and Community-Based System Dynamics Process, in: Community Based System Dynamics. Springer, pp. 17–30. https://doi.org/10.1007/978-1-4614-8763-0_2

Hovmand, P.S., Andersen, D.F., Rouwette, E., Richardson, G.P., Rux, K., Calhoun, A., 2012. Group Model-Building 'Scripts' as a Collaborative Planning Tool. Systems Research and Behavioral Science 29, 179–193. https://doi.org/10.1002/sres.2105

Kahneman, D., 2011. Thinking, fast and slow. macmillan.

Maani, E., Kambiz, Cavana, Y., Robert, 2007. Systems thinking, system dynamics : managing change and complexity. Prentice Hall, Auckland, N.Z.

Meadows, D., 2008. Thinking in systems. Chelsea Green Publishing.

Nabavi, E., 2022. Computing and Modeling After COVID-19: More Responsible, Less Technical. IEEE Transactions on Technology and Society 3, 252–261. https://doi.org/10.1109/TTS.2022.3218738

Scott, R.J., Cavana, R.Y., Cameron, D., 2013. Evaluating immediate and long-term impacts of qualitative group model building workshops on participants' mental models. System Dynamics Review 29, 216–236. https://doi.org/10.1002/sdr.1505

Scriptapedia Wikibooks contributors, n.d. Scriptapedia. https://en.wikibooks.org/wiki/Scriptapedia (accessed 3.8.23).

Sterman, J., 2000. Business dynamics. Irwin-McGraw-Hill.

Stilgoe, J., Owen, R., Macnaghten, P., 2013. Developing a framework for responsible innovation. Research Policy 42, 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Vennix, J., 1996. Group Model Building. Wiley, New York.