If this, then that, then what? Beyond the archetypes: a generative process for building responsible causal models

C. A. Browne[a] **and E Nabavi**[a]

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

Abstract:

The need to consider effective techniques for bridging qualitative and quantitative SD approaches was raised by experts in the field as a potential area of improvement to broaden the adoption of SD in higher education institutions. Of particular concern was the challenges and opportunities arising from introducing system archetypes to novice modellers in the context of different disciplines and application spaces. We do not take a position on the validity of either argument, and instead examine how we can reframe the relationships between the system archetypes as a diagnostic tool to become a generative tool for qualitative or quantitative models, and in turn encourage responsible modelling practices in a qualitative setting. Built on a simple scaffold: "If this, then that, then what?", a *many models are better than one model* position is taken. An experiment was conducted with 30 novice modellers to explore the potential of such a tool, and in the space of approximately 45 minutes, one causal link led 30 participants to generate 131 causal hypotheses, which were categorised into 34 groups, and could be used as the basis for a generative set of models. Further work is required to explore the full potential of this approach, but in practice, we have found that this process is a simple scaffold for non-expert modellers to develop an understanding of fundamental feedback structures, promote diversity of thought within a group setting, and overcome issues arising from teaching system archetypes, enabling even novice modellers to explore hindsight about past behaviors and dynamic hypotheses about our future.

Keywords: *Systems thinking, qualitative modelling, system archetypes.*

1. INTRODUCTION

This work responds to conference prompt *How do models help us build both hindsight about past behaviors and dynamic hypotheses about our future?* by proposing a novel approach in a well-discussed area of qualitative system dynamics (SD) modelling: the system archetypes. This paper proposes a methodology to overcome issues with using system archetypes with novice modellers, described as the 'If this, then that, then what?' approach.

In section 1, we outline the broader background and motivation for this topic with respect to the concept of responsibility in qualitative system dynamics (SD) modelling. In section 2 we describe the "If this, then that, then what?" approach in general terms, and in section 3 we report on the methodology for an initial experiment that utilises the approach. In section 4 we report on initial results, and in section 5 raise points of discussion, which leads to the conclusions in section 6.

1.1. Background and Motivation

The intent of the paper is not to further arguments that endorse or critique the value of the system archetypes, but to provide a pragmatic methodology that helps novice modellers engage with small models that display dynamic complexity so that they can gain insight about a given situation.

The motivation for this work arose in two phases: first, in discussions within the Asia-Oceania region during 2020 about the disestablishment of a number of critical SD university-level courses in the region due to staff and budget cuts in the wake of COVID-19, and second with university-level educators and practitioners as part of a strategic set of workshops hosted by the System Dynamics Society (the University Innovation project) during 2021-23 that aims to support action that helps disseminate SD methodology in higher education institutions to support the growth of the field.

Of the many topics of conversations engaging veteran and novice educators were the topics of curricula and the processes by which these curricula are taught. Put simply and without further analysis or judgement, it would be a fair representation to say that the field could be divided into two: those that use the system archetypes as an entry point into the field, and those that do not. Further work would be required to examine whether factors such as year level, prerequisite knowledge, disciplinary application, support for SD, or other factors led to this apparent bifurcation of views.

Eliciting the arguments for and against either approach is not the purpose of this paper—although an updated critical treatment of the counter-intuitive system archetypes, canonical situation models and abstracted micro-structures of Lane and Smart (1996) given the widespread popularity of the system archetypes (Kim, 1992; Meadows, 2008; Senge, 2006) in the intervening decades would help inform these discussions—there appears to be a counter-intuitive balancing narrative at the core of each side's approach:

Premise:	We have limited space in curricula to teach SD	
	For quantitative approaches	*For qualitative approaches*
Therefore:	We need to give students a strong theoretical foundation	We need to give students useful take-aways
But:	There is a high overhead cost to teach quantitative approaches	There is a limit to the usefulness of qualitative approaches
Lament:	Students don't have the qualitative skills to translate insights into actions	Students don't have the quantitative skills to build insights from simulations
And:	Only the student understands the model	The student is anchored to the archetype

This, of course, is not universal, and there are many excellent examples where either approach can be successfully implemented. But, for many educators, these problems are real, and methodology to help bridge the gap between qualitative and quantitative SD is needed. If we take a step back from these discussions, it is obvious to see that the root problem is not necessarily with either approach, but rather that university curricula are fragmented and often crowded, space is limited and time with students is precious. Further, it was observed in these discussions that these bases provide an unstable platform for further work, such as in doctoral research.

The leads to the motivation for this work: we need simple scaffolds that can be picked up by novice modellers quickly, encourage good SD habits of thought to develop, and allow for the unbounded simulation creativity that SD modelling can provide. In this context, the applicability of this work is primarily aimed at those that engage in qualitative SD modelling, such as those who may be using the system archetypes and as a protocol for participatory modelling, which often operates under similar constraints of time as in the classroom.

1.2. Responsible Qualitative SD Modelling

The notion of responsible modelling has been developing quickly in the digital age, and arises from gaps in areas such as machine learning and artificial intelligence, which are often described as black-box models. As a recent example, the role of models in decision making has been highlighted during the policy chaos of COVID-19, where modellers of all stripes were attempting to help decision-makers with foresight but were often based on poor assumptions or incomplete data (Jalali et al., 2020; Spalluto et al., 2020). Various governments, research organisations and areas of industry have been calling for development of protocols that can enable greater trust in models through transparency, explainability, repeatability, removing bias, and addressing ethical concerns (Fjeld et al., 2020; Gunning et al., 2019; Lu et al., 2022; Nabavi and Browne, 2023; Vyhmeister et al., 2023).

The field of simulation SD models based on stock-and-flow structures have been ahead of this trend, with well-defined approaches for model conceptualisation through the modelling process (Sterman, 2000, pp. 83–105), and protocols for model transparency and repeatability (Martinez-Moyano, 2012; Rahmandad and Sterman, 2012). However, qualitative SD models are fuzzier due to the incompleteness of mental models , are often not intended to be simulated and rely on narrative structures to explain (Maani and Cavana, 2007; Meadows, 2008; Sterman, 2000)., making them open to interpretation and less repeatable.

One area of qualitative SD practice that has been advancing approaches for model conceptualisation is the broad field of participatory modelling (Andersen et al., 1997; Hovmand, 2014; Vennix, 1996), where 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 (Andersen and Richardson, 1997; Hovmand et al., 2012; Scott et al., 2013; Scriptapedia Wikibooks contributors, n.d.) These protocols can help build in notions of transparency and repeatability. However, there are still challenges that are such as capturing different and shared causal reasoning between multiple stakeholders, and blending these perspectives in a useful and responsible way.

Time factors, the fuzzy nature of qualitative modelling, the social nature of group processes, 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 qualitative modelling practices.

The area of system archetypes is further problematic. These often seductively insightful stories shortcut the wisdom that is gained by really grappling with the unique dynamics of a given situation: they show the simplicity that lies on the other side of working through complexity. As a result, it is easy for a novice participant to become blinkered, and go around fitting situations to the archetypes, rather than developing the skills involved in building a model that represents any given situation.

To this point, we take the position that small models can deliver valuable insights in group settings (Newell, 2012), and approach building the notion of responsibility into system archetypes by extending George Box’s oft-quoted aphorism *all models are* wrong; some are useful by adding many models are better. The concept of plurality in many models sits well in a participatory setting, where a large part of the challenge is integrating the many different perspectives and world views, and the struggle to escape to the bounded rationality that arises with the convergent thinking observed in the system archetypes.

1.3. What Reverse engineering the system archetypes might teach us

The system archetypes have been widely popularized in recent decades (Kim, 1992; Meadows, 2008; Senge, 2006; Wolstenholme, 2004). The predominant framing of the archetypes has been as a diagnostic tool (Goodman and Kleiner, 1993), simplifying complex dynamic situations into well-trodden problem-solution structures (Braun, 2002), making it easy for decision-makers to conceptualize and gain a level of insight into a problem. As a diagnostic tool, the system archetypes family tree allows easy framing and diagnosis of problem-solutions (Braun, 2002; Goodman and Kleiner, 1993; Senge, 1997).

It is from here that the approach for developing a simple scaffold to gain insight through small models arises. If we examine the structure of the system archetypes family tree, there becomes apparent an underlying structure of combinations of one-four Balancing and Reinforcing loop structure. An example of this reverse engineering of the diagnostic family tree to become a generative family tree is shown in Figure 1.

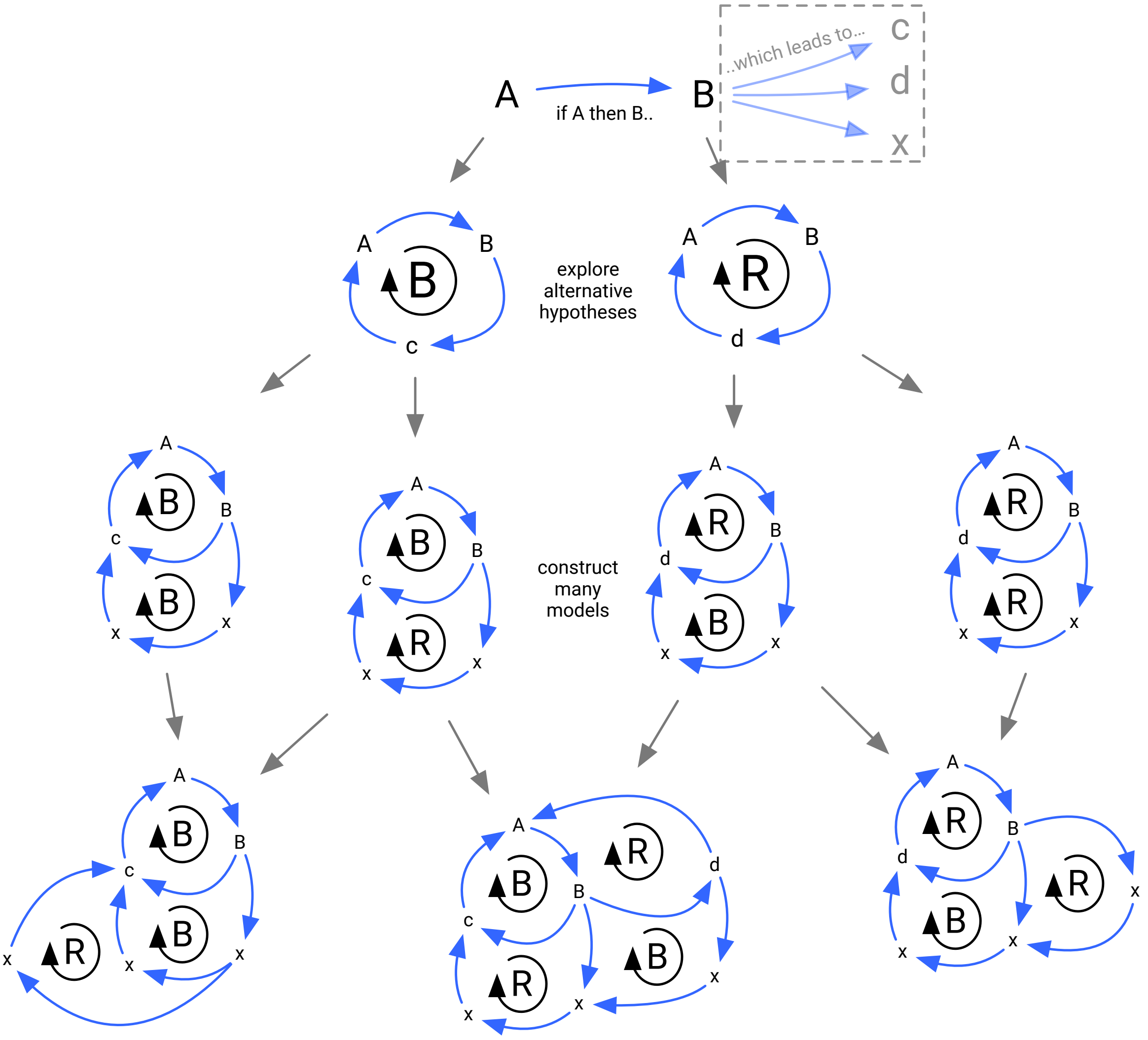

Figure 1: Generative Family Tree structure

Rather than a diagnostic structure, this becomes a generative structure for developing incrementally complex models, move from a 'single model' approach to the plurality of a 'many models' approach, and is a simple scaffold to build up the capacity of novice participants to develop their own insights about a given problem.

2. APPROACH

To help overcome some of the implicit biases that can occur by using the system archetypes, we have developed a simple generative methodology to help participants provide a base to explore many alternative dynamic 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, creative, and designed overcome the bounded rationality arising from the system archetypes.

The main process involves four steps:

Step 1: If this, then that...
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.

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: …then what?
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:

- an outcome that might make things better
- an outcome that might make things about the same
- an outcome that might make things worse

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.

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: Look for dominant feedback.

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. Find the loops that are likely to dominate behaviour. These loops become the basis for a broader discussion in Step 4.

Step 4: Discuss and explore combinations.

In Step 2, many overlapping causal links will be generated. Use the alternative links to foster discussions within groups and broaden out the family tree. If time 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 that comes from the diagnostic frame of system archetypes. 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 1. 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 1: Overview of timing used in the experiment (total time approximately 45 min)

Time	Description	Actor/s
5 min	Introduction to systems thinking	Researchers
5 min	Overview of experiment and Human Ethics protocol	Researchers
5 min	Introduction to problem statement and sharing of news video	Researchers
3 min	Explanation of step 1, then time for participants to complete step 1 by answering the prompt in the first empty box	Individual participants
5 min	Explanation of step 2, then time for participants to complete step 2 individually by answering the prompts in the next three empty boxes	Individual participants
2 min	Explanation of feedback behaviour	Researchers
5 min	Explanation of step 3, then time for participants to complete step 3 by identifying system behaviour and perceived likelihood	Individual participants
10 min	General discussion between groups about different answers. Participants encouraged to write any notes	Small groups
5 min	Concluding remarks	Researchers

Participants completed an individual worksheet shown in Figure 2 throughout the experiment, which were collected and used to generate the data for analysis. The textual descriptions provided in the free-text responses were categorized into the substantive themes for the analysis in §4.

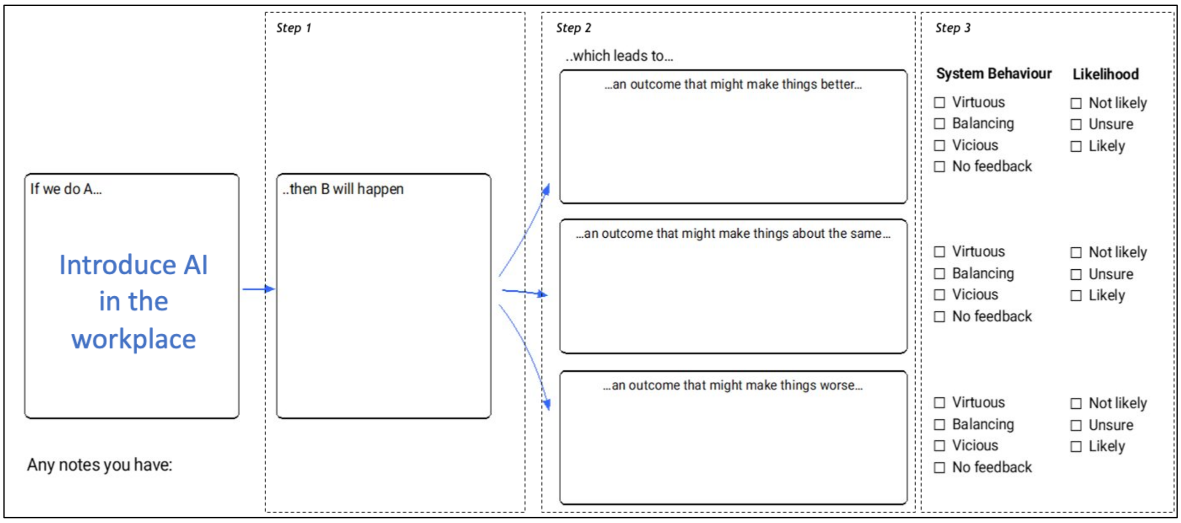

Figure 2: The snapshot of worksheet used to structure the activity and capture data. Note: ‘steps’ indicated by the dotted boxes represent the steps shown in Table 1, and were not included on the worksheet.

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 categorized 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 2.

Table 2: 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” “A lot of tasks will be automated, things like translation, editing, writing emails”
1,2	efficiency	concerning the speed or time to undertake a task	“Make work more efficient” “Increased performance efficiency”
1,2	employment	concerning jobs or replacing humans	“Some people will lose their jobs (replaced by AI)” “May cause some sort of job cut & unemployment”
1,2	productivity	concerning volume or processes of output	“Improve the productivity in general” “[...]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” “Customers are not satisfied by AI […] services.”
2	errors	concerning creation or identification of errors	“Mistakes made by AI are sometimes are inevitable” “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.” “Increase profit”
2	outputs	concerning the output or process itself	“AI doesn’t improve the product itself.” “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” “Open opportunities for work/life balance”

As described in §1.1, many educators and participatory model facilitators are limited in time and scope to effectively engage participants in model building. This workshop is no exception, required to fit into the schedule of a one-off one-hour class. Nonetheless, the generation of the causal insights in the worksheets and the overall discussion can still be applied back into the generative family tree in a post-hoc asynchronous way. This is often the case in participatory models, where the expert modeler navigates the data generated through qualitative responses to generate a model that represents the situation. The results from the worksheet are shown in §4, and subsequent discussion of the generative family tree in §5.

4. RESULTS

30 participants recorded responses during the workshop. 9 responses in step 1 had multiple responses, where more than one statement was provided. The 39 relationships described in step 1 resulted in 131 causal chains, when step 2 multiple responses were included. In some cases, no response was captured, which has been reported as ‘nil response’. A total of 34 relationships between step 1 and step 2 were drawn between categories in step 1 and 2. A summary of descriptive results by workshop step is shown in Table 3.

Table 3: Summary of results by workshop step

Step	Response	Count	(%)
	Total	**39**	-
	employment	13	(33)

Step	Response	Count	(%)
2 ‘then what’	Total	**131**	-
	better	46	(35)

1 'then this' free text	automation	11	(28)
	efficiency	10	(26)
	productivity	3	(8)
	quality	2	(5)
2 'then what' free text	**Total**	**131**	-
	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)
	nil response	2	(2)

direction of behaviour	same	43	(33)
	worse	42	(32)
3 system behaviour	Total	**131**	-
	virtuous	45	(34)
	balancing	30	(23)
	vicious	34	(26)
	no feedback	9	(7)
	nil response	13	(10)
3 likelihood	**Total**	**131**	-
	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 3 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.

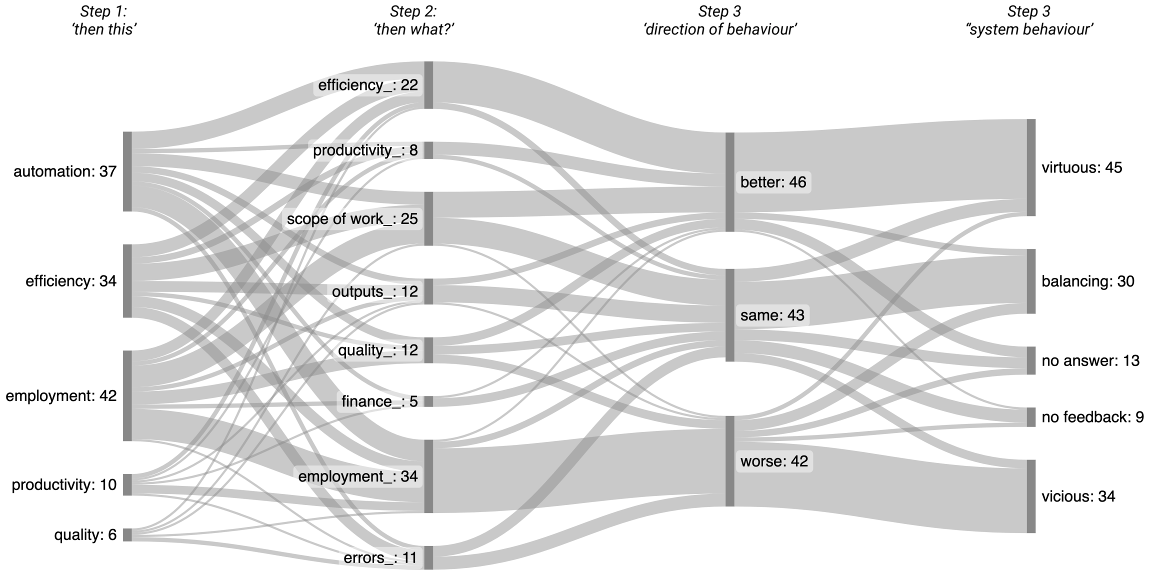

Figure 3: Visualisation of responses by workshop step

Section 5 will examine the collected causal links to develop a generative family tree that represents the results collected in the workshop.

5. DISCUSSION

The relationships shown in Figure 3 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.

During the workshop, there were some noticeable trends in Figure 3 that represent the collective mental models of the participants that could be used to prompt or prime discussion prior to commencing a formal model. Working from right to left in Figure 3:

> *Q: What trends are there between the direction of behaviour and the system behaviour?*
>
> Comment: The 'better' responses tend to exhibit 'virtuous' behaviour; the 'worse' responses tend to exhibit 'vicious' behaviour. It is worth noting that both the vicious and virtuous behaviours will have reinforcing feedback structures, and that these loops may indeed be working against each other. The 'same' responses tend to exhibit 'balancing' behaviour.
>
> *Q: What trends are there between our 'then whats' and the direction of behavior.*
>
> Comment: 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'*
>
> Comment: 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.

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

For the purposes of this exploration, we will not explore the construction of the small causal loop or stock-and-flow diagrams, but rather look for the likely dominant loops and their behavior to demonstrate the concept. In this step we have identified the dominant themes, and thought about what we might call that loop based on the sentiment in the workshop outputs:

- R1 [Rise of the robots]: Automation, employment and efficiency
- B1 [Same job, different work]: Automation, employment and scope of work
- R2 [Errors everywhere]: Automation, efficiency and errors
- B2 [Robot's servants]: Automation, efficiency and scope of work

These loops then become a starting point for the exploration of the generative family tree, and provide a strong shared narrative for the inclusion of experience, validation with historical narratives, sense-checking from different perspectives, discussion of further dynamic hypotheses, and the construction of simulation models. Such a conceptual generative family tree using the hypothesized loop descriptions above is shown in Figure 4. Although it is represented as a series of simplified causal loops, there is no reason that the same process could not be used to construct as a series of incremental stock-and-flow models.

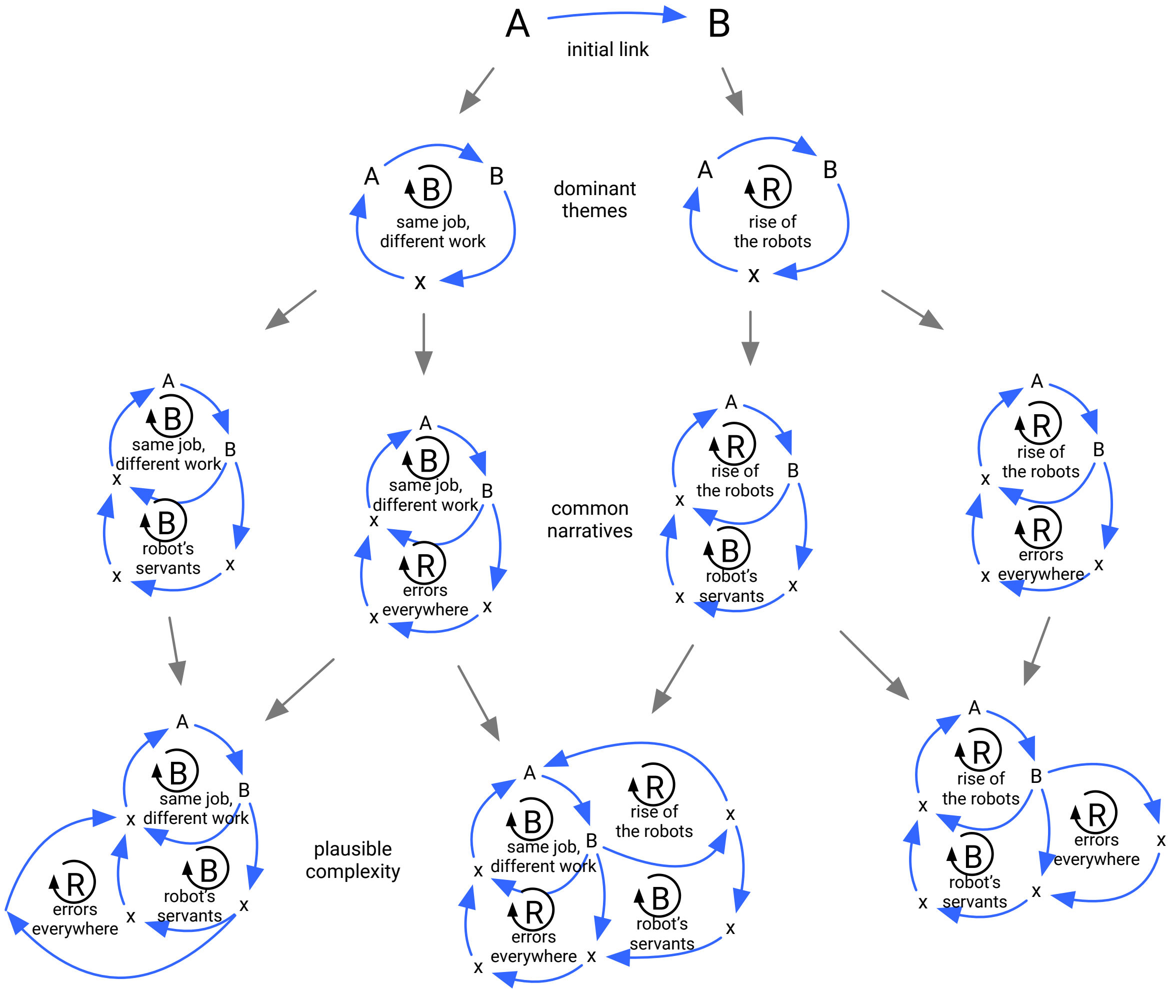

Figure 4: Hypothesized generative family tree. Although represented as a series of simplified causal loops, there is no reason that the same process could not be used to construct as a series of incremental stock-and-flow models.

6. CONCLUSION

The need to consider effective techniques for bridging qualitative and quantitative SD approaches was raised by experts in the field as a potential area of improvement to broaden the adoption of SD in higher education institutions. In this paper we have described a process developed to help facilitate the construction of models in a time constrained setting, such as in an introductory system dynamics class or a participatory model building activity. This qualitative SD modelling process arises from the considerable body of work that examines the system archetypes, which has been discussed as a problematic topic, especially for novice modellers. The 'If this, then that, then what?' structure is a simple scaffold that encourages participants to explore multiple hypotheses in the problem space, and can be then applied to generate many models on a given topic.

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, and could be used as the basis for a generative set of models. 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'. 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 introductory system dynamics classes, or participatory modelling processes. Further work is required to explore the effectiveness of this technique beyond the scope of the experiment set at the initial phase.

In practice, we have found that this process is a simple scaffold for non-expert modellers to develop an understanding of fundamental feedback structures, promote diversity of thought within a group setting, and overcome issues arising from using system archetypes with novice modellers.

7. 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., sWells, 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

Braun, W., 2002. The system archetypes. System 2002, 27.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M., 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. SSRN Journal. https://doi.org/10.2139/ssrn.3518482

Goodman, M., Kleiner, A., 1993. Using the archetype family tree as a diagnostic tool. The Systems Thinker 4, 5–6.

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.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.-Z., 2019. XAI-Explainable artificial intelligence. Sci Robot 4, eaay7120. https://doi.org/10.1126/scirobotics.aay7120

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

Jalali, M.S., DiGennaro, C., Sridhar, D., 2020. Transparency assessment of COVID-19 models. The Lancet Global Health 8, e1459–e1460. https://doi.org/10.1016/S2214-109X(20)30447-2

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

Kim, D.H., 1992. Systems Archetypes I. Pegasus Communications.

Lane, D.C., Smart, C., 1996. Reinterpreting 'generic structure': Evolution, application and limitations of a concept. System Dynamics Review 12, 87–120. https://doi.org/10.1002/(SICI)1099-1727(199622)12:2<87::AID-SDR98>3.0.CO;2-S

Lu, Q., Zhu, L., Xu, X., Whittle, J., 2022. Responsible-AI-by-Design: a Pattern Collection for Designing Responsible AI Systems. https://doi.org/10.48550/arXiv.2203.00905

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

Martinez-Moyano, I.J., 2012. Documentation for model transparency. System Dynamics Review 28, 199–208. https://doi.org/10.1002/sdr.1471

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

Nabavi, E., Browne, C., 2023. Leverage zones in Responsible AI: towards a systems thinking conceptualization. Humanit Soc Sci Commun 10, 1–9. https://doi.org/10.1057/s41599-023-01579-0

Newell, B., 2012. Simple models, powerful ideas: Towards effective integrative practice. Global Environmental Change 22, 776–783.

Rahmandad, H., Sterman, J.D., 2012. Reporting guidelines for simulation-based research in social sciences. MIT web domain.

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 [WWW Document]. URL https://en.wikibooks.org/wiki/Scriptapedia (accessed 3.8.23).

Senge, P.M., 2006. The fifth discipline : the art and practice of the learning organization. Doubleday/Currency, New York.

Senge, P.M., 1997. The fifth discipline fieldbook. Random House Digital, Inc.

Spalluto, L.B., Planz, V.B., Stokes, L.S., Pierce, R., Aronoff, D.M., McPheeters, M.L., Omary, R.A., 2020. Transparency and Trust During the Coronavirus Disease 2019 (COVID-19) Pandemic. Journal of the American College of Radiology 17, 909–912. https://doi.org/10.1016/j.jacr.2020.04.026

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.

Vyhmeister, E., Castane, G., Östberg, P.-O., Thevenin, S., 2023. A responsible AI framework: pipeline contextualisation. AI Ethics 3, 175–197. https://doi.org/10.1007/s43681-022-00154-8

Wolstenholme, E., 2004. Using generic system archetypes to support thinking and modelling. System Dynamics Review 20, 341–356. https://doi.org/10.1002/sdr.302