Online Appendix

Appendix 1 System Dynamics Modeling

System dynamics (SD) is useful for policy analysis and design for problems arising in complex social, managerial, economic, or ecological systems (Forrester 1961, Richardson and Pugh 1981, Sterman 2000). Qualitative diagrams and simulation models are tailored to the specific problem under review. Discussions with system stakeholders are used to understand the structure of the system, and to develop a qualitative diagram (in the form of a Causal Loop diagram). This diagram visualizes important variables and their relationships. The diagram is then transformed into a system of mathematical relationships between variables. This system of equations is referred to as a simulation model. Available datasets, literature review and stakeholders' knowledge are used to identify parameter estimates for key variables and to ensure that the simulation model is a reasonable reflection of the real system. The simulation model is used for policy formulation and analysis. It also provides a dynamic hypothesis bringing to light the feedback processes resulting in the projected behaviors. This dynamic hypothesis can be communicated in the form of a Causal Loop Diagram and in the form of a simulation model.

SD is different from other simulation modeling approaches because it takes a holistic view of all organizations and processes involved in the system, incorporates feedback loops and dynamic processes, and includes nonlinearities in the relationships between variables (Andersen, Rich et al. 2020). SD modelers work extensively with key stakeholders and experts to develop the structure of the system and incorporate data from numerous sources (Vennix 1996). Although all simulation models are imperfect reflections of reality, working closely with stakeholders throughout the process can increase the simulation model's accuracy and legitimacy.

The model is tested by comparing model output to empirical data, and if discrepancies exist, refining the model and parameter estimates. Once the model has been developed and tested, inputs can be modified to conduct "what if" analyses of how short- and long-term outcomes would change in response to various policy scenarios. In addition, the process of developing the model and running the base case scenario may expose new concepts and previously unknown but significant variables (Richardson 2013). For example, combining numerical data, written data, and the knowledge of experts in mathematical form may identify inconsistencies about how we think the system is structured and how it behaves over time (Forrester 1975).

SD modeling can therefore be seen as progressing linearly through stages of development, testing, and analysis. Model testing and analysis create many opportunities for finding discrepancies between a model and case study data sources, including stakeholder understanding(Andersen, Luna-Reyes et al. 2012, Tomoaia-Cotisel, Allen et al. 2022). Discrepancies may also be discovered by comparing to modeler expectations and literature. Numerous formal and informal 'tests' describe useful opportunities to discover different types of discrepancies(Lane 1995, Barlas 1996, Forrester and Senge 1996). Acting on these opportunities creates a circular modeling process which goes from development to testing to analysis and continually returns back to development when making a correction, thus starting the process over again(Saeed 1992, Martinez-Moyano and Richardson 2013). These model refinement cycles continue until they are constrained by research resources including time, data, money, research team members' skill, and stakeholders' expectations(Groesser and Schwaninger 2012).

SD models do not generate a forecast or claim that an outcome will have a specific value at some future point in time. Rather, SD models predict dynamic implications of policies to determine whether they will result in a future that will be better or worse than it would have been without the intervention. The primary output of a model is a set of graphs over time illustrating how key variables will change in the future under different scenarios. Systems modelers seek to understand how internal policies, internal decisions, and external phenomena interact to generate the problems observed over time. An explicit goal of SD is to provide an explanation for *why* and *how* the outcome will change, potential unintended consequences, and areas where implementation may not lead to intended outcomes.

The SD approach is best suited for problems that are dynamic (they involve change over time), have feedback (the transmission and return of information), nonlinearities, and delays between cause and effect (Forrester 1961, Sterman 2000). This approach has been used to address economic, social, management, energy, supply chain, and problems in the medical sciences (Barlas 2002).

Appendix 2 Orientation to the CLD

In this section, we orient the reader to how a CLD is read. The links between the arrows identify the direction of causality and loop identifiers indicate if the loop is reinforcing or balancing (R and B). To read the causal links between variables you can start with the variable at the tail and, as a heuristic, ask if the variable at the tail increases, which direction does the variable at the head (the arrowhead) move. If an increase in the variable at the tail causes the variable at the arrowhead to increase, then it gets a plus sign and captures that it is moving in the same direction. If an increase in the variable at the tail caused the variable at the arrowhead to decrease, it would be labeled with a minus sign. This indicates the direction of causality. For example, if the Staff QI Effort was increased then Operations Capability would move in the same direction and the arrow would be labeled with a plus sign. In turn, an increase in Operations Capability will cause Variances and All-Cause Harms to decrease. The minus sign indicates that the variable at the arrowhead moves in the opposite direction. Logically, if Staff Operations Capabilities increases you would anticipate that All cause Harms would decline.

These causal connections indicate the relationship between the variables one at a time. They eventually feedback upon themselves and form a feedback loop. Here we walk through one reinforcing loop and one balancing loop to orient the reader to how loops are read.

In the R3 feedback loop in Figure 2, the Staff QI Effort influences staff UBA Capability which in turn influences Motivation for QI and back to Staff QI Effort which closes the feedback loop. This loop was labeled the **R3 UBA Practice Makes Perfect** loop and captures the effect of a change in any variable in the loop. The reinforcing nature of this loop (labeled R3) means that any initial change will work its way around the loop to reinforce the direction of initial movement. An increase in staff UBA Capabilities provides Motivation for QI, which in turn, increases the Staff QI Effort and thus staff UBA Capabilities. Likewise, all else equal, a decrease in UBA Capabilities decreases Motivation for QI, which, in turn, decreases further Staff QI Effort and feeds back to reduce further UBA Capability. Once moving, and with all other things held constant, this feedback loop will continue to perpetuate itself in a vicious or a virtuous cycle depending on the initial direction of the change. It will do so until another of the interconnected feedback loops in Figure 2 becomes dominate and changes the behavior.

The B3 feedback loop is one of the loops that eventually works to control the R3 feedback loop. When a quality objective (difference between Goal Variances and actual variances) departs from the desired range, it creates Goal Pressure which motivates staff to use QI. This Staff QI Effort increases staff Operations Capability which in turn reduces Variances and All-cause Harms which closes the balancing feedback loop. This loop was labeled the **B3 Working Smarter** loop and captures the effect of a change in any variable in the loop. The balancing nature of the loop means that any initial change will work its way around the loop to control the R3 feedback process.

Appendix 3 Comparing UNDAUNTED with MUSIQ and NPT versions

Here, we present the results of our comparison of UNDAUNTED CLD with CLD-like versions of MUSIQ and NPT. This comparison is inspired by mental model comparison methods employed in SD. Significant interpretation was required to convert MUSIQ and NPT to CLD form, making it less useful to employ a formal comparison method.

This appendix presents causal loop diagrams (CLDs). Gray links indicate ambiguous link polarity. Arrows are labeled using the French for with *avec* and without *sans*, to reference semantic differences in structure. Sans indicates that the link uses implicit mechanisms to describe issues made more explicit in the UNDAUNTED CLD. In other words, it may be oversimplifying the relationship. For instance, in the upper left of **Figure 9**, the link comment (Sans "Pressure for Brain Drain", "Staff") indicates that the link connecting *Microsystem Capability* directly to *QI Team Decision-making Process, Team Norms & Skills* may imply two variables as mechanisms. This is because UNDAUNTED has "Pressure for Brain Drain" and "Staff" as explicit mechanisms in the causal chain between semantically-similar variables. *Avec* indicates that the comparison CLD may be overcomplicating a relationship, whether that is in UNDAUNTED or not. In figures comparing a theory CLD with UNDAUNTED, blue boxes show differences in boundaries.

All comparison models are reported as having ambiguous link polarity. No attempt is made by these authors to describe the elements which are distinctive to CLDs: delays, link polarity, and loop polarity. As such, these elements are not included in the causal loop versions. Loop labels are added in one situation where numerous loops are labeled and then reoriented for overlaying with UNDAUNTED.

Mapping MUSIQ versions indicates a tendency to oversimplify operational aspects of QI. Mapping NPT versions indicates a tendency to simultaneously oversimplify and overcomplicate key issues in implementation in a service operations context. The most common way of oversimplifying things is to draw links that seem to imply many intermediate mechanisms. Overcomplication comes in two forms: duplicate variables and duplicate causal pathways. This comparison indicates that the boundaries of MUSIQ and NPT versions are generally smaller than UNDAUNTED and that they consider causality and feedback in a limited manner.

Overall, the comparisons of CLD details and boundaries highlight deficiencies in both MUSIQ and NPT relating to their lack of holism and nuance in capturing the complexities and dynamics of quality improvement and implementation within a healthcare service delivery context.

Figure 9 illustrates the causal loops implied in MUSIQ version 1. It shows 1 link with *avec* and 4 links with *sans* comments. This comparison indicates a tendency in MUSIQ version 1 to oversimplify aspects of healthcare operations such as brain drain, human resources, staff learning, errors, the closed-loop of QI, and the time and effort required for QI. It also suggest a link between QI capability and operations capability, which was also speculatively suggested in our study.

Figure 9: MUSIQ version 1 causal loops

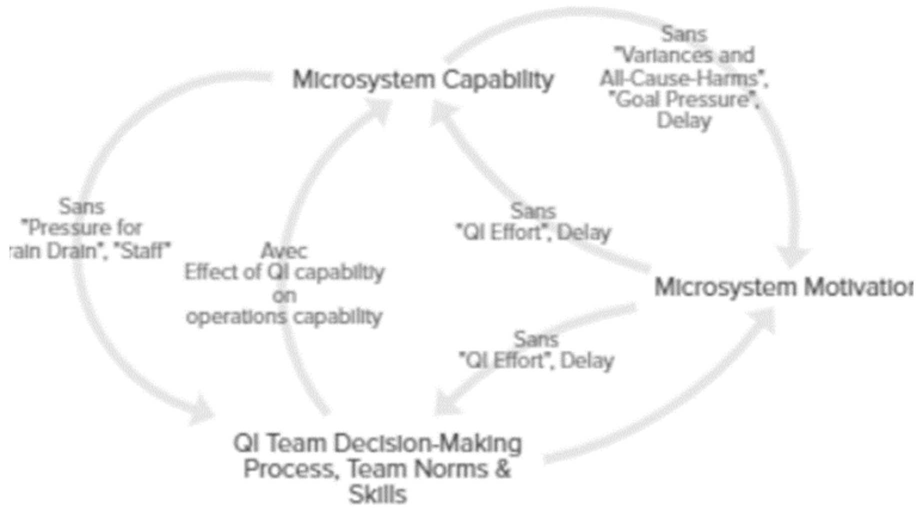


Figure 10 compares the causal loops in MUSIQ version 1 with the UNDAUNTED CLD. In the overlay, it appears that reducing the differences in boundary would mean expanding MUSIQ version 1 to consider causality and feedback more explicitly and to consider the causal importance of healthcare operations, as noted above.

Figure 10: MUSIQ version 1 causal loops and UNDAUNTED

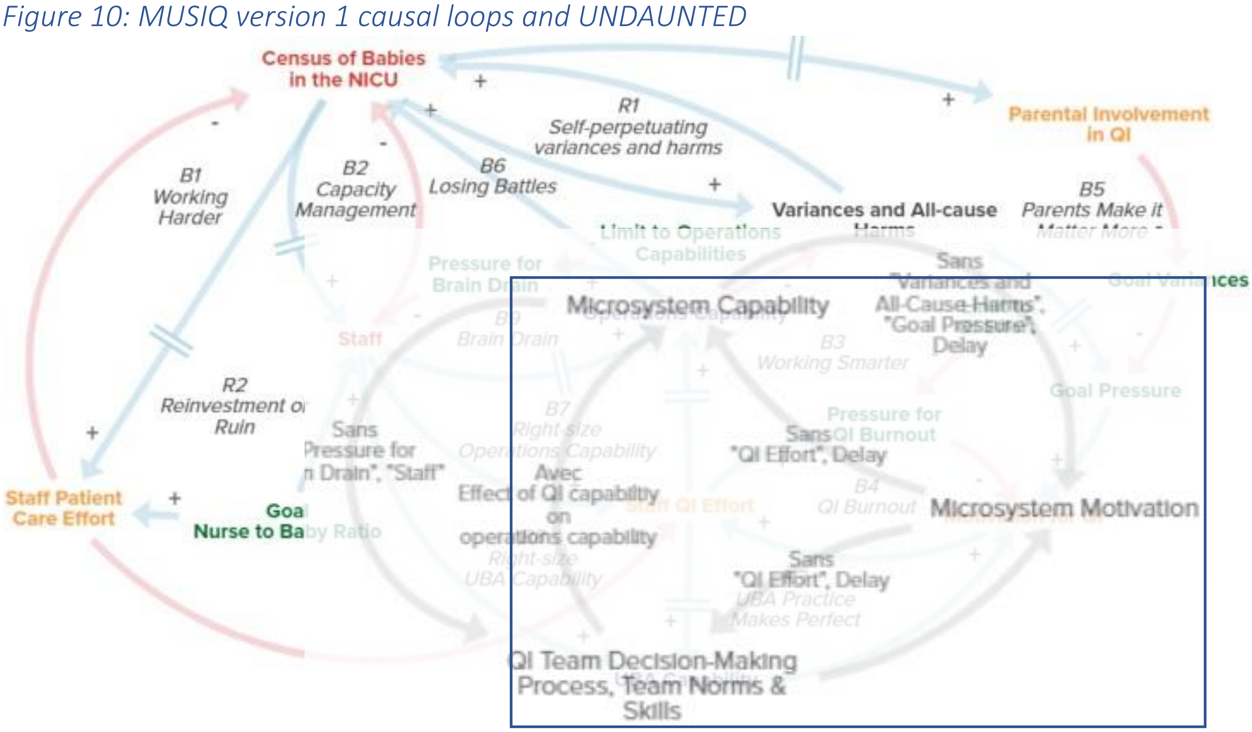


Figure 11 illustrates the causal loops implied in MUSIQ version 2. It shows 1 sans link and 2 *avec* links. This indicates that MUSIQ version 2 continues oversimplifying operations. It also suggests possible additional organizational issues to explore in future research on UNDAUNTED.

Figure 11: MUSIQ version 2 causal loops

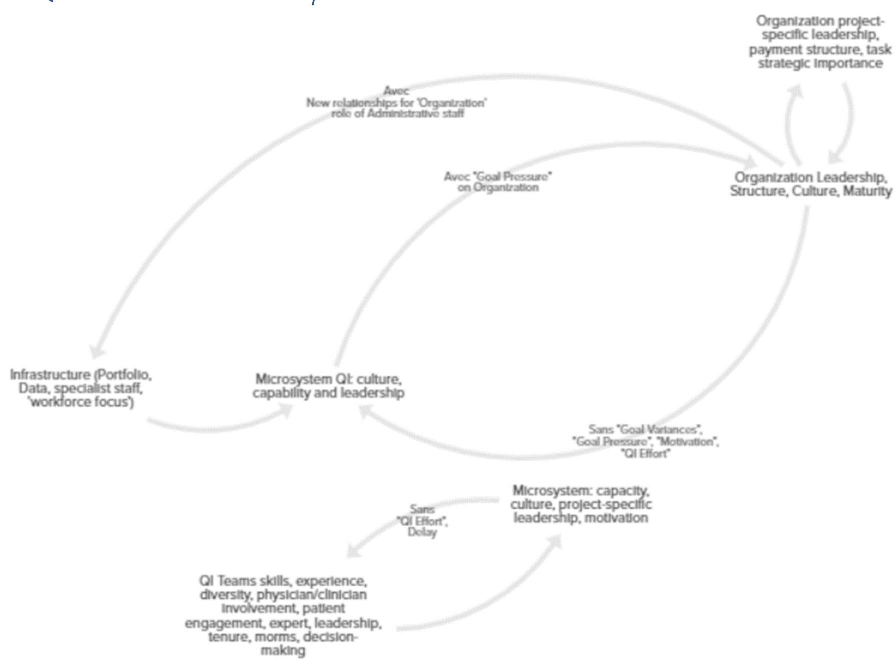


Figure 12 compares the causal loops identified in MUSIQ version 2 with the UNDAUNTED CLD. The overlay shows that reducing the differences in boundary would mean changes to both.

Figure 12: MUSIQ version 2 and UNDAUNTED

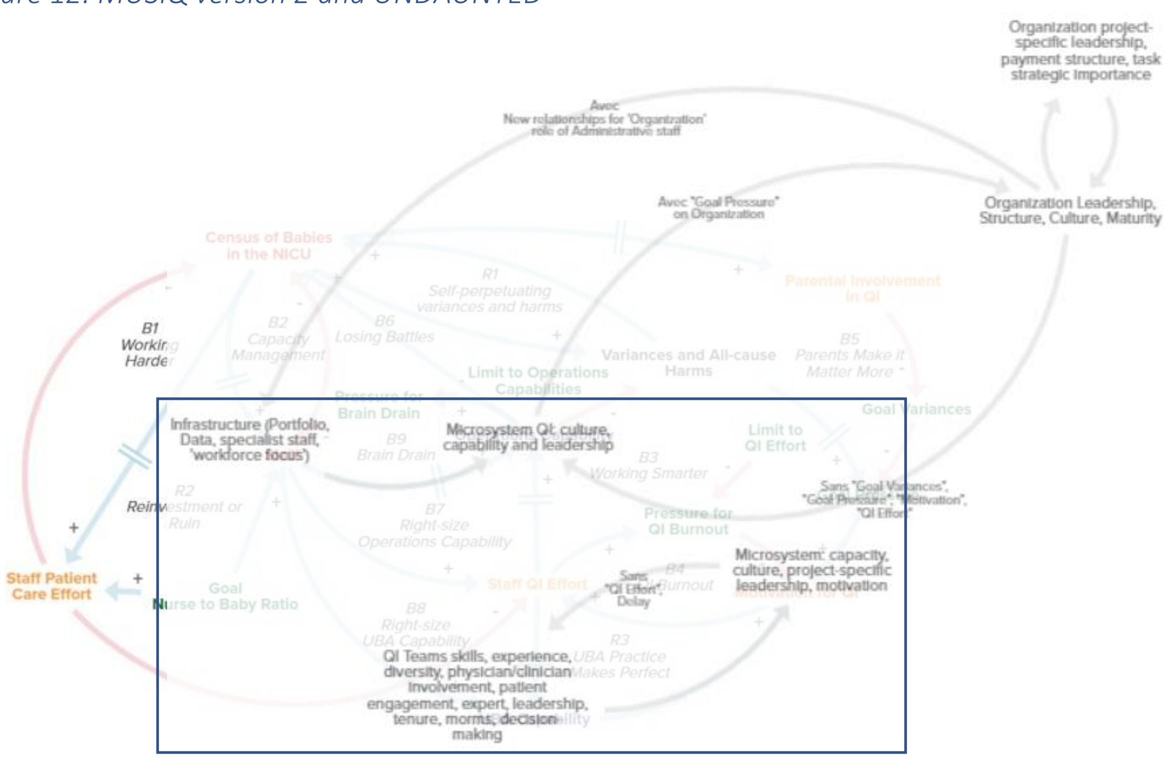


Figure 13 shows the causal loops identified in the Normalization Process Model, a precursor to Normalization Process Theory (NPT). It shows 2 sans links. This indicates Normalization Process Model oversimplifies motivation and the closed-loop of QI.

Figure 13: Normalization Process Model causal loops

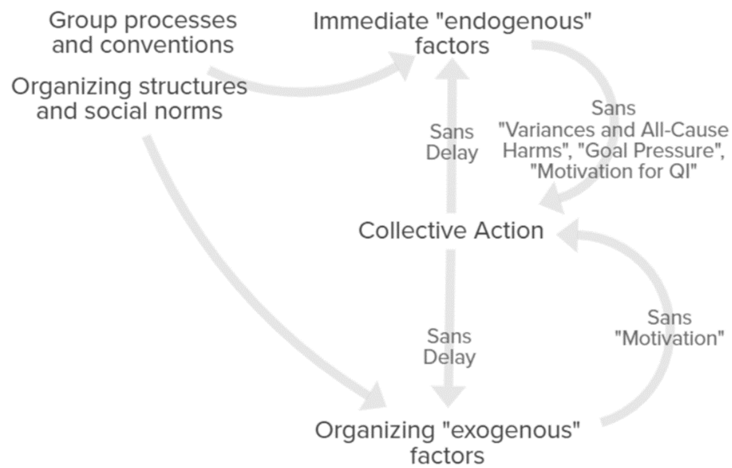


Figure 14 illustrates the causal loops identified in the Normalization Process Model, a precursor to Normalization Process Theory (NPT). It shows 2 sans links. This indicates Normalization Process Model oversimplifies motivation and the closed-loop of QI and overcomplicates the role of human resources.

Figure 14: Normalization Process Model and UNDAUNTED

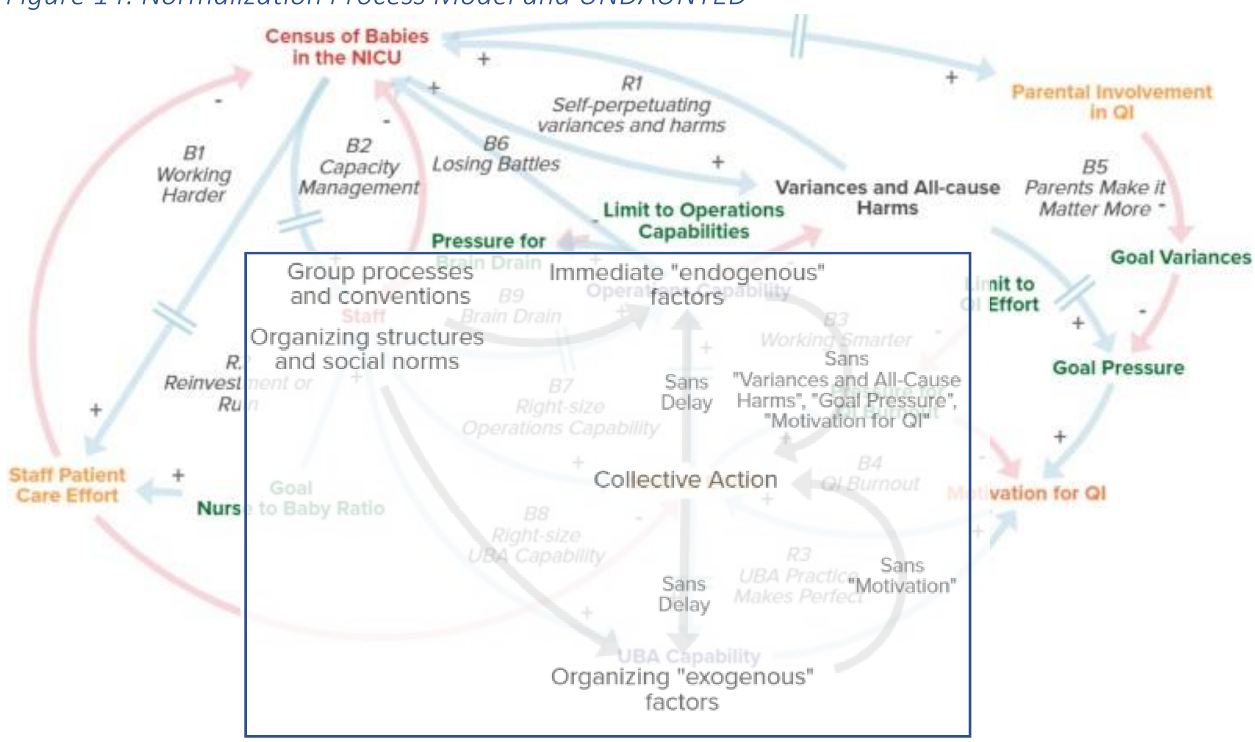


Figure 15 displays the causal loops implied in NPT version 1. It shows 1 link with *avec* and 9 links with *sans* comments. This comparison indicates a tendency in NPT version 1 to oversimplify links of healthcare operations such as human resources, staff learning delays, errors, and goal-seeking. It also suggests a link between motivation and goal pressure which relates to making a consensus commitment, which we assigned to the link between QI capability and motivation.

Figure 15: NPT version 1 causal loops

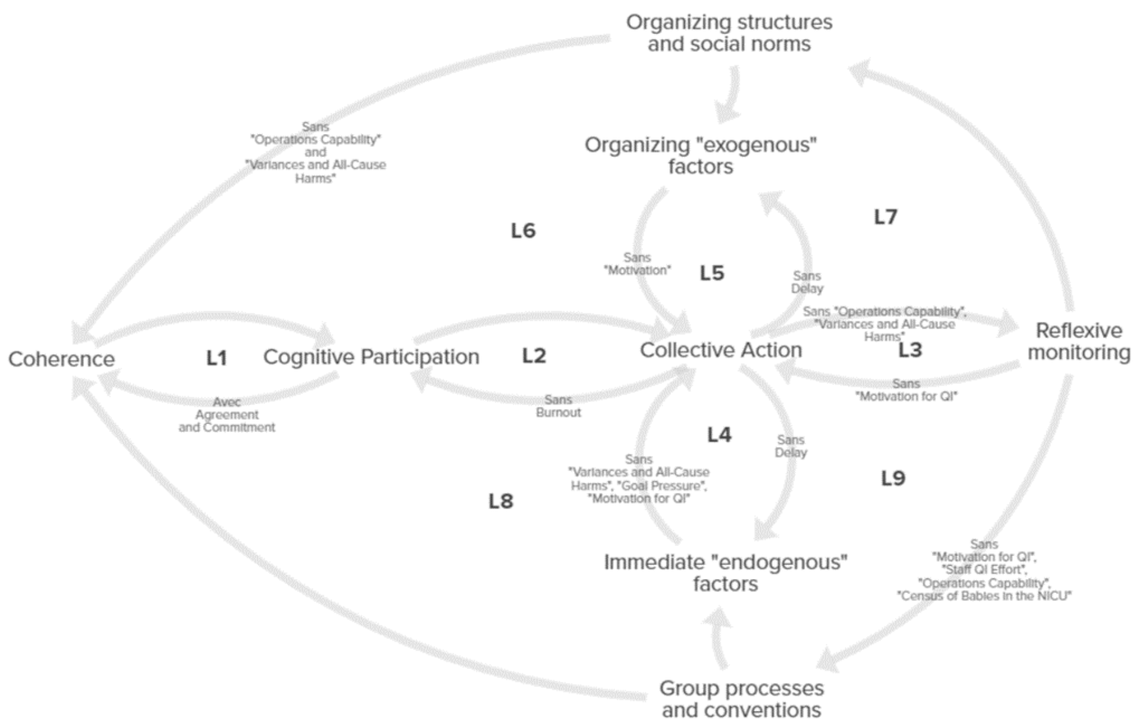


Figure 16 displays the causal loops implied in NPT version 1, reoriented based on the overlay with UNDAUNTED. The main difference with Figure 15 is that two variables are overlapping. The overlapping variables on the left show NPT version 1's tendency to under-appreciate how human resources practices shape group norms and the ones on the right show its tendency to under-appreciate how goal-seeking behavior relates to the understanding and meaning people derive from improvement. The upper links show its tendency to link variables in ways that duplicate its own existing pathways. Overall, it tends to overcomplicate things.

Figure 16: NPT version 1 causal loops (reoriented)

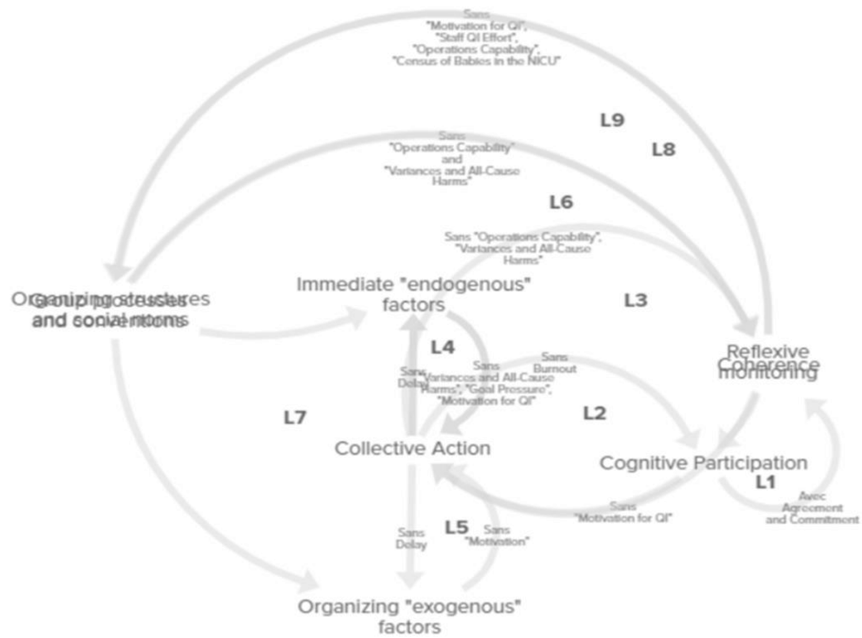


Figure 17 compares the causal loops identified in NPT version 1 with the UNDAUNTED CLD. The overlay shows that matching the boundary would mean redefining and enlarging NPT version 1.

Figure 17: NPT version 1 causal loops (reoriented) and UNDAUNTED

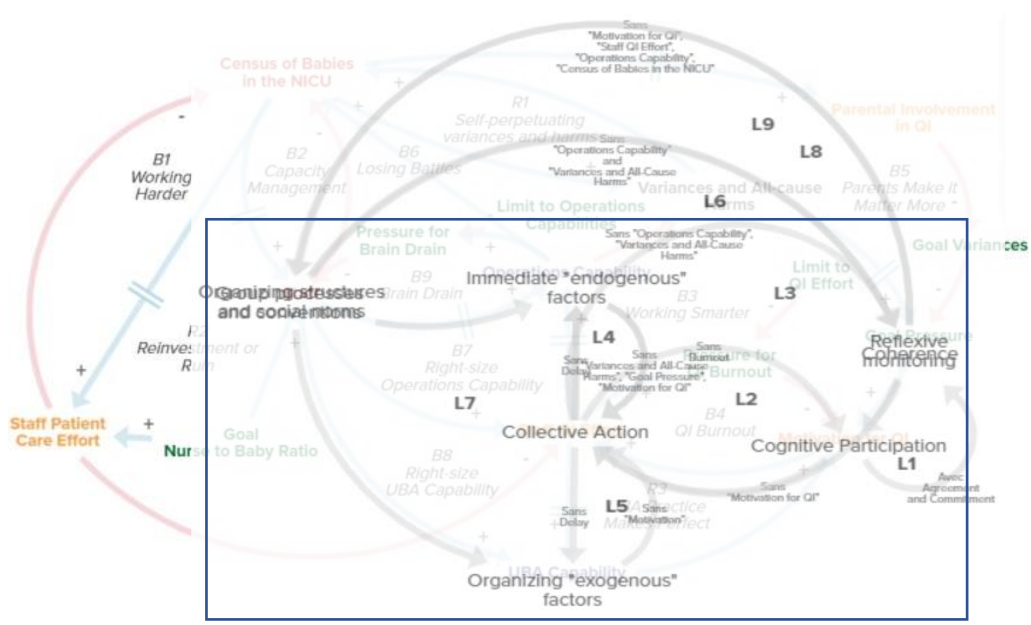


Figure 18 displays the causal loops implied in NPT version 2. It shows 5 links with *sans* comments. This comparison indicates a tendency in NPT version 2 to oversimplify healthcare improvement by ignoring delays and downplaying the role of motivation.

Figure 18: NPT version 2 causal loops

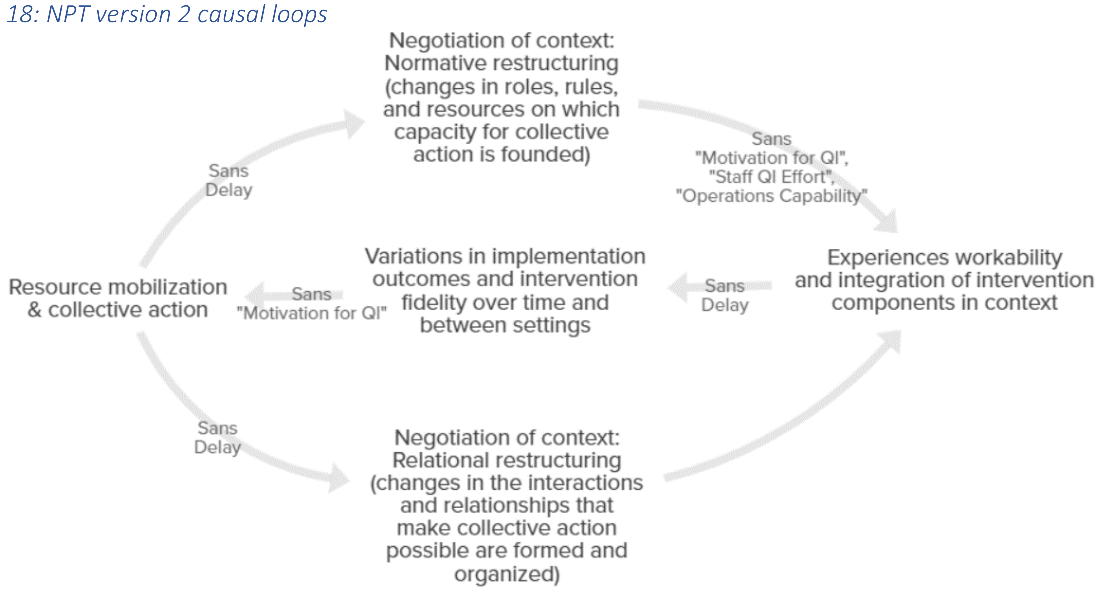
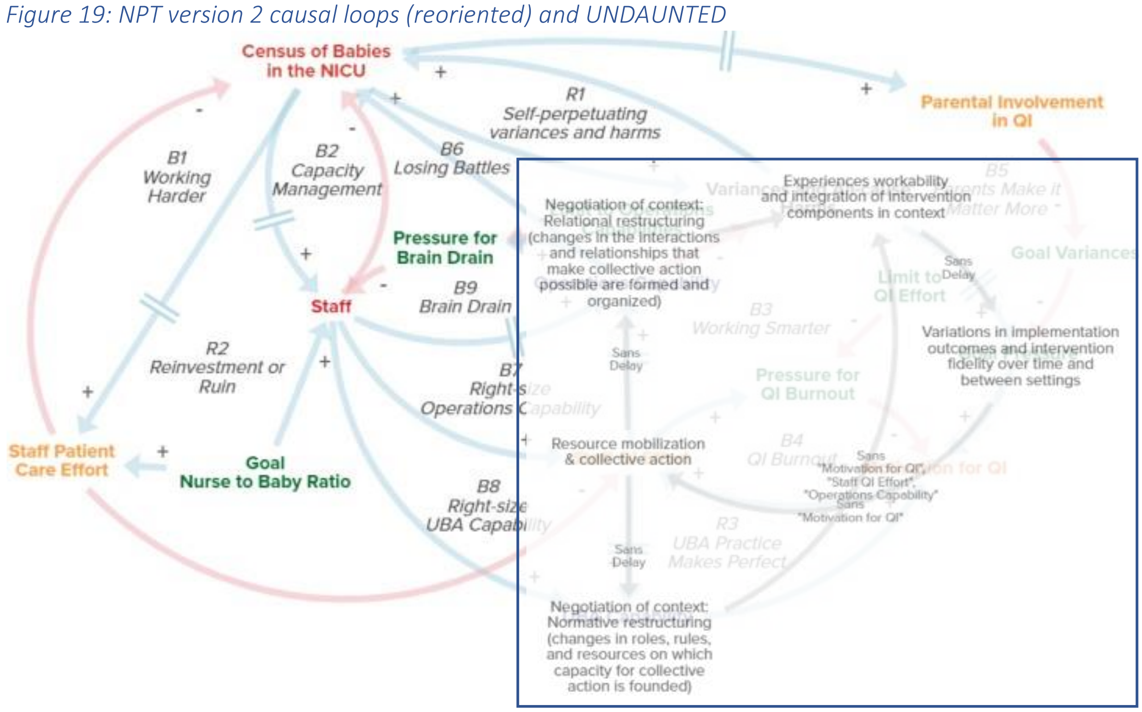


Figure 19 compares the causal loops identified in NPT version 2 with the UNDAUNTED CLD. The overlay shows that matching the boundary would mean simplifying variable meanings and enlarging the scope of NPT version 2 to include feedbacks with operations. Notably, the tendency to overcomplicate and duplicate continues, largely by removing motivation.

Figure 19: NPT version 2 causal loops (reoriented) and UNDAUNTED



Appendix 4 Model Documentation

Model documentation is available on request.