Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo and M. Loper, eds.

MANAGING THE COMPLEXITY OF COLLABORATIVE SIMULATION-BASED HUMAN-IN-THE-LOOP EXPERIMENTATION

David Prochnow
Robert Portigue

Simulation, Experimentation, and Gaming Department
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102, USA

ABSTRACT

Human-in-the-Loop (HITL) experimentation can be effective for assessing the efficacy of multi-person operations. Based on the experiment objectives of such operations, an experiment team creates a virtual experimentation environment in which multiple human subjects can collaborate to achieve mission objectives. Afterwards, analysts assess the effectiveness of new technologies or procedures under test. While there is much value in conducting simulation-based HITL experimentation, there is also a large degree of complexity. This paper presents a framework for managing the complexity of executing such an experiment by dividing the experiment team into several smaller specialized teams that collaborate using processes described in this paper. An innovation leadership team, scenario team, technical team, and Data Collection and Analysis (DCA) team work together to plan and execute the experiment and assess results. An experiment team can use the methodology presented here to manage complexity and, ultimately, accomplish the objectives of collaborative simulation-based HITL experiments.

1 INTRODUCTION

This paper provides guidance on collaborative HITL simulation-based experimentation with the goal of reducing complexity to facilitate the accomplishment of experiment objectives. These types of experiments involve multiple human subjects who work together to achieve a stated objective. Experiments may assess the operations of an emergency response team, military operations, law enforcement actions, or any mission that requires the collaboration of multiple persons. These persons may come from different organizations, each of which has its own procedures and chain of command. Simulations provide a virtual environment in which the human subjects operate. In some cases, humans may interact directly with a simulation, while in other cases humans may employ a real or emulated operator system (e.g., a military command and control systems, air traffic control system, etc.) that interfaces with the simulated world. The human participants also communicate with each other using standard communications devices, such as phone and chat. This type of experiment is suited for assessing the best way for humans to employ a new concept, which may be a new or improved system, a new technology that is applied to multiple systems, or a new methodology.

This paper discusses how best to conduct an individual experiment, based on over 10 years of conducting these types of experiments at MITRE's National Security Experimentation Laboratory (NSEL) for multiple United States government organizations, mainly the Department of Defense, but also including numerous non-defense entities as well. In this paper we will discuss the following:

- Complexities of HITL Simulations (Section 2)
- Process for planning, executing, and analyzing HITL experiments (Section 3)
- Examples of Simulation-Based HITL Experiments (Section 4)

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

- Summary (Section 5)

Processes are described in this paper, with key points made based on lessons learned from over 60 experiments. Many of these points are also enforced by other seminal experimentation reports, including (Alberts and Hayes 2002), (The Technical Cooperation Program 2006), and (Harvey et al. 2003).

Note: While it is often useful to conduct a campaign of experiments, that is not the subject of this paper. More information about experiment campaigns can be found using the sources in the "References" and "Additional Reading" sections.

2 COMPLEXITY OF SIMULATION-BASED HITL EXPERIMENTS

Conducting a successful HITL experiment is a complex endeavor. This includes project complexity and simulation complexity. For project complexity, we use the definition provided by Girmscheid and Brockmann: "the degree of manifoldness, interrelatedness, and consequential impact of a decision field" (Girmscheid and Brockmann 2008). The experiment environment is typically a System of Systems (SoS), using the United States DoD definition of SoS as "a set or arrangement of systems that results when independent and useful systems are integrated into a larger system that delivers unique capabilities" (Department of Defense 2004). The SoS may include newly developed simulations, legacy simulations, and special purpose software, all of which exchange data at runtime. In addition, humans interact with the experiment infrastructure, so the unpredictability of human behavior is also a factor. As we are discussing simulation-based experiments, simulation complexity is also a key issue, and it involves both structural complexity and software complexity. As stated by Popovics and Monostri, structural complexity involves the number of objects and attributes to be represented while software complexity refers to programming logic and use of computing resources (Popovics and Monostri 2016).

Alberts and Hayes, in their *Code of Best Practices for Experimentation* document (Alberts and Hayes 2002), describe three major phases of an experiment, as summarized in Table 1. As Alberts and Hayes emphasize, the most important phase of the experiment is the pre-experiment phase. Each experiment must have a focus on the objectives and the questions the experiment is seeking to answer, and what data is to be collected and analyzed to achieve the experiment objectives. The upfront planning will affect how the experiment is designed, what type of experiment scenario is developed, how the technical infrastructure is constructed, and the requirements for data collection and analysis. This technical infrastructure must represent a scenario that achieves the objectives. Of course, the experiment is only useful if appropriate data is collected and analyzed to determine the merits of the concept being examined, so both the scenario and the technical infrastructure must be geared towards generating and capturing appropriate metrics. Managing the complexity of the experiment depends heavily on the pre-experiment phase.

The experiment execution phase must also be managed carefully. During the execution phase, the scenario and technical infrastructure must be resilient enough to handle unexpected actions by the participants, and the experiment team must be prepared to deal with unexpected situations.

The post-experiment phase, in which data is analyzed, is also very important, but it is much more straightforward than the other phases. While this phase requires hard work, it is lower complexity than the pre-experiment and experiment execution phases.

The full experiment process requires a set of persons with different skill sets. This includes persons knowledgeable of the concept to be tested, persons to create an appropriate scenario, a technical team to create the simulation environment, and analysts to collect and assess the data generated in the experiment. Team members with different roles must coordinate closely with one another, or the experiment is doomed for failure. For example, if the experiment objectives are not articulated well, the rest of the experiment team will not be able to build an appropriate experiment environment. In addition, the scenario team cannot devise a scenario that the technical team cannot implement, the technical team cannot develop and integrate systems that do not produce the data needed by the analysts, and the analyst team cannot build a Data Collection and Analysis Plan without understanding the scenario and technical components under development. The complexity of coordinating different aspects of these HITL experiments must be managed to achieve a successful outcome.

Table 1: Summary of Experiment Phases, from Alberts and Hayes (2002).

Phase	Sub-Phase	Description
Pre-Experiment	Formulate the experiment	Define objectives and hypothesis. Make explicit assumptions. Determine what the experiment will assess. Identify key variable relationships.
	Establish the experiment team	Build a team with appropriate skill sets. Divide the team into sub-teams for the innovation leadership team, scenario team, DCA team, and technical team.
	Prepare the initial experiment plan	Determine, in general, how experiment will be conducted, what key variables will be modified and measured, how data will be collected, and how it will be analyzed. Develop rough experiment plan.
	Create the detailed experiment plan	Develop scenario, create a Design of Experiments (DOE), determine technical components to be developed or integrated, create Data Collection and Analysis Plan (DCAP), and develop the detailed experiment plan.
Experiment Execution	Set up experimentation environment (rehearsal setup)	Develop and integrate components for the simulation-based technical infrastructure, refine scenario, test systems and system interfaces.
	Conduct a dry run	Train persons who will participate in dry run, conduct full end-to-end run, assess any issues revealed in testing, determine action to take prior to experiment conduct.
	Execute the experiment plan	Provide training to participants, conduct the experiment runs with close monitoring to identify any issues, and collect data.
Post-Experiment	Conduct data analysis	Analyze the results from quantitative and qualitative sources.
	Interpret the data	Determine the significance of the data in context with data from other experiments and research.
	Distribute the results	Draft the experiment report, collect feedback, develop final report, and distribute the results.

3 CONDUCTING HITL EXPERIMENTS

This section builds upon the process documented by Alberts and Hayes in their *Code of Best Practices for Experimentation* document (Alberts and Hayes 2002). This paper describes how to divide the experimentation team into multiple specialized teams that must work together. These four teams are:

- **Innovation Leadership Team.** Formulates the experiment and provides managerial and programmatic leadership through the entire process.
- **Scenario Team.** Develops the contexts in which the experiment will be conducted.
- **Data Collection and Analysis Team.** Determines what data needs to be generated and captured, and how that data will be assessed.
- **Technical Team.** Develops the infrastructure on which the experiment will be executed.

These teams each play critical roles in the experiment planning, execution, and analysis processes. In addition, these teams must collaborate to ensure (1) the objectives are well understood by all, (2) the scenario is appropriate for the new concept to be assessed, (3) there is a plan for how data will be collected and analyzed, and (3) the technical infrastructure and technology to be utilized by operators provides a suitable representation of the scenario and allows collection of the data for analysis.

The following subsections discuss each of the experiment phases, outlining the tasks for each of the teams described above. Process diagrams for each phase illustrate the methodology, and the accompanying descriptions emphasize key points for each phase. (These diagrams do not currently use any formal modeling specifications, such as Business Process Modeling Notation (BPMN), but this is under consideration.)

The subsections below assume a complex experiment that involve over 20 human subjects, several federated simulations, and multiple types of operator systems. For simpler experiments, some of the steps may not be needed. In addition, for less complex experiments, work may be allocated to a person or persons that represent two or more of the teams shown above.

3.1 Pre-Experiment Phase: Experiment Formulation

This initial experiment planning phase, if done well, will provide an initial understanding of the issues to be addressed, the scope of the experiment, and an initial model of the interaction of relevant variables. This phase just uses a small core of team members, and additional members are recruited in the following phase.

During this phase, the innovation leadership team must explicitly define objectives and hypotheses (Harvey et al. 2003), as the rest of the experiment will evolve from this point to meet the stated objectives and assess the hypotheses. In addition, the team should make explicit assumptions to limit the scope of the experiment. These assumptions must be clearly documented; otherwise, there is a risk that experiment results will later be misinterpreted if the assumptions are neglected.

During formulation, the innovation leadership team should develop a general idea of what the experiment will assess. That is, the team must determine the independent variables (what will be varied in each run), the dependent variables (what are the outputs that will be potentially affected by the independent variables), and what are the control variables (what will be held constant each run). The key variables relationships should be well understood, ideally by developing diagrams for which there is consensus before proceeding further.

3.2 Pre-Experiment Phase: Establishment of Team

After the experiment is formulated, the Innovation Leadership Team must build a team to plan, execute, and analyze the experiment. As indicated previously, the experiment team is typically divided into four sub-teams consisting of an innovation leadership team, scenario team, DCA team, and a technical team. These are the people who will ensure that the experiment meets the stated objectives.

For the scenario team, appropriate subject matter experts should be recruited to design the experiment and support development of the scenario that will be portrayed in the experiment. The scenario team should understand both the new concept being explored and the context in which it will be used.

Assembling a data collection and analysis team requires finding people who will objectively and skillfully collect the needed data and perform the subsequent analyses. It is important that the members of this team are free of any conflict of interest that would question their analysis.

The technical team will develop, integrate, and test the experiment environment in collaboration with the scenario developers and analysts. There will be some give-and-take with the scenario and analysis teams, so in addition to be technically adept, the technical team must also have the social skills to interact with the other teams.

3.3 Pre-Experiment Phase: Initial Experiment Planning

After the experiment has been formulated, the experiment team must start determining how the experiment will be conducted, including treatment of relevant variables, how data will be collected, and how it will be analyzed. The ultimate output of this phase is a rough experiment plan. Figure 1 shows the processes for the initial planning phase.

During this phase, the scenario team should revisit the variable relationships established during experiment formulation. They should ensure that the chosen dependent variables are reliable indicators of the idea to be assessed. In addition, they should select enough dependent variables to obtain valid experiment results. The scenario must also ensure that the set of independent variables is sufficient for determining whether the concept can be adequately assessed. The team must also determine how independent variables will be introduced in the experiment. The independent variables could be modified by using simulation configuration settings, simulating planned events, injecting events from an experiment control team, or mandating that human participants employ different processes each run.

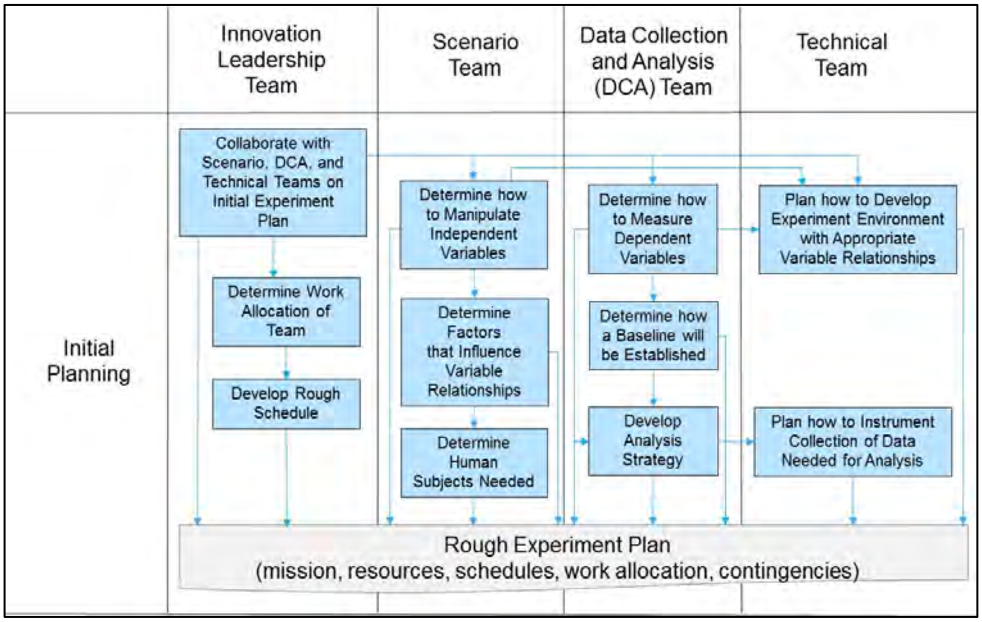

Figure 1: Initial Experiment Planning Process.

The scenario team should also devise an initial scenario that is credible, but to reduce complexity, it should not be any more detailed than is needed. The scenario must also be feasible and provide realism for operators within the given experiment resources. This requires working with the technical team to ensure they can build an experiment environment to adequately represent the scenario.

Another role of the scenario team at this stage is to determine what categories of human subjects will be necessary. Ideally, the recruited participants will be end users of the innovation under test. The human subjects will also need to be trained, so the scenario team should begin to consider how the participants will be trained and what resources will be required for doing so. The participants must have adequate time to train, as the new concept cannot be assessed to its fullest without ensuring proficiency of the users.

The data collection and analysis team must determine how the dependent variables will be measured. This requires working with the technical team to determine what can be automatically captured using the simulation infrastructure. For data that cannot be automatically captured, the DCA team may identify other means, such as the use of human observers, post-run surveys for the participants, or after-action reviews. The DCA team should also determine how a baseline will be established, in order to compare the "to be" idea with the "as is." This could be accomplished by using one of the simulation runs to derive a baseline, or by using known knowledge. Alternatively, threshold and objective criteria can be set for the experiment data that is to be collected.

The technical team works with the scenario team and the DCA team during this phase. This team determines how it will develop the experiment environment based on the envisioned scenario and what data needs to be collected. The technical team needs to be clear about what is feasible, and feedback from the technical team may result in changes to the scenario or the DCA plan.

The entire experiment team should use the initial planning period to identify actions that will require long lead times, such as facility reservations and certain hardware procurements.

The ultimate outcome of the initial planning phase is a rough experiment plan. This will identify the major milestones, with work allocated among the technical, scenario, and DCA teams. The milestones should ensure adequate time for integrating and testing the experiment infrastructure and processes.

3.4 Pre-Experiment Phase: Detailed Experiment Planning

After an initial plan has been documented, the experimentation team must then work out the details for a more comprehensive plan. Figure 2 shows the process diagram for this phase, leveraging the rough experiment plan developed during initial planning. During this phase, the experiment team evolves the rough experiment plan with details from which the experiment can be created.

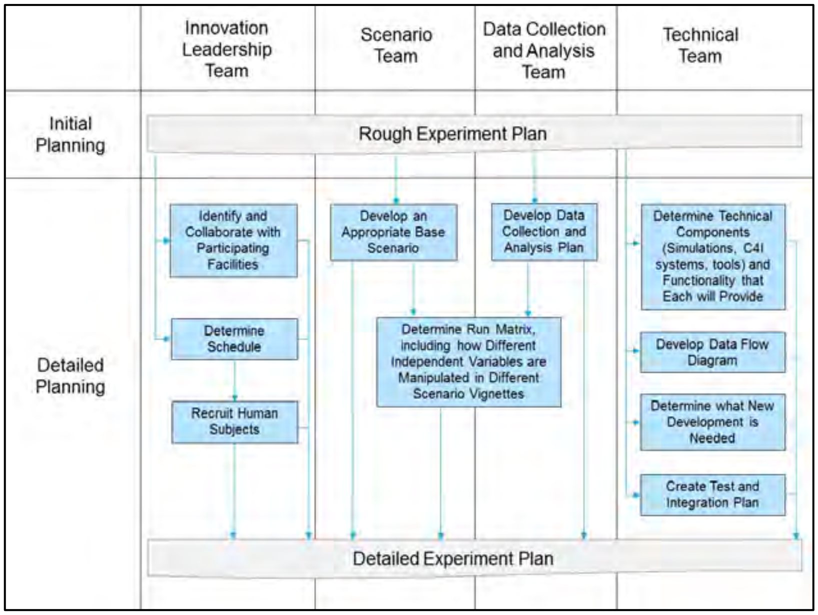

Figure 2: Detailed Planning Processes.

The scenario team must finalize the scenario during this phase. They should work with subject matter experts in developing key aspects of the scenario. If the scenario lacks credibility, the experiment findings might be discounted or discarded altogether. An ideal scenario will allow participants to respond creatively to experiment events. That is, it should not be overly scripted so that the innovation under test can be explored in a dynamic environment.

The DCA team must develop a Data Collection and Analysis Plan (DCAP). This decomposes the experiment objectives into measurable components (i.e., the dependent variables). There should not be any dependent variables that have no relation to the experiment objectives. For each metric, they determine how the data will be captured. This requires coordination with the technical team on what can be automatically collected using the simulation infrastructure. The DCA team also determines any data reduction and other processing needed to put collected data into a form in which it can be analyzed. Development of a comprehensive DCAP is a key principle of (The Technical Cooperation Program 2006).

The scenario and DCA teams work together to determine a run matrix that dictates how different variables are manipulated in different scenario vignettes. The run matrix is informed by the base scenario and DCAP developed by the scenario team and DCA team, respectively. As the number of runs for a HITL experiment is limited by the duration in which the humans are available (unlike a non-HITL experiment that may employ a simulation to run many iterations in a batch mode on a high-performance computer), there should be a good idea of what will be learned from each run.

The technical team identifies the specific components and associated functionality that will be used in the experiment. This may be a combination of existing systems and systems that require creation or modification to meet the experiment requirements. The technical team also determines how all components will interoperate, devising a data flow diagram. An integration and test plan is also devised during this stage.

The Innovation Leadership Team also plays some key roles during this phase. They identify a master schedule for managing the experiment progress. They also collaborate with participating facilities, ensuring that each experiment facility has appropriate availability, not just for the experiment itself, but for testing and setup beforehand. This team also recruits the human subjects, using the participant requirements developed during initial planning, and identifying specific persons. The recruited persons should be appropriate for the role they will play, and they should also have open minds that are not attached to old technologies or processes. The participants should be available for the entire experiment; otherwise, results will be skewed by the different capabilities of human subjects that take turns in the same role. It should be conveyed to persons being recruited that the focus of the experiment is the concept being assessed; if participants think they are being graded, that may lead to risk-averse behavior. Care should be taken to choose participants who are comfortable thinking outside the box in an experimental environment. They should be made to feel that they are part of the experiment. This will allow for innovative solutions to problems as they occur.

At the end of this phase, the team should have a detailed experiment plan that guides the remaining experiment development, conduct, and analysis.

3.5 Experiment Execution Phase: Rehearsal Setup

Once the detailed planning has occurred, the experiment team can focus on setting up the infrastructure needed for conducting the experiment. This consists of the development, integration, and testing of system components. Figure 3 depicts the rehearsal setup processes.

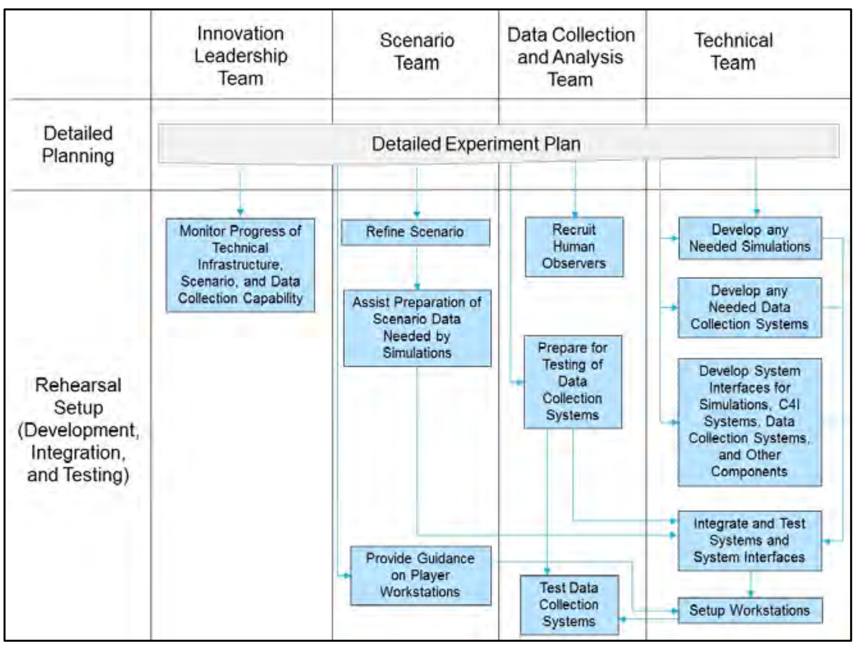

Figure 3: Rehearsal Setup Processes.

This phase requires obtaining existing simulations, as well as the development or modification of any simulations, as specified in the detailed experiment plan. For experiments that leverage multiple simulations, the technical team must determine which parts of the scenario are represented by each simulation. Collaboration with the scenario team can lead to the determination of how and where each element of the scenario is being represented. In some cases, scenario events may not be represented in software, but instead be orchestrated from the Experient Control (sometimes referred to as the “White Cell”).

Simulations will require appropriate data, so the sources of this data should be identified early. This may require the technical team working with subject matter experts to obtain appropriate data sources. Data will also have to be converted into formats that the simulations need. The technical team should not discount the time and effort to convert raw data into formats that are usable by different simulations, each of which may require the data in different formats, different units, and at different levels of aggregation.

In addition to the simulations, the technical team must obtain or develop the systems that represent what the operators will used in the real world. This may involve integrating the actual real-world system if it can be obtained and integrated into the simulation environment, or it may require developing an emulation of the real-world system. If a system must be emulated, then the developers must work with subject matter experts to ensure the system is represented adequately for the experiment.

The technical team must also determine and document the data that needs to be exchanged between systems. To avoid unnecessary software development and thereby reduce complexity, the technical team should leverage existing middleware and interoperability standards to the extent possible. Development of new middleware is often necessary for interfacing systems that do not normally exchange data.

Integration events should be conducted early and often. Early events may focus on pairwise testing of two systems, while latter events include more of the systems functioning together. All components should be tested in the experiment environment.

3.6 Experiment Execution Phase: Rehearsals

Prior to actual execution, the experiment team should conduct one or more full rehearsals (i.e., dry runs) to ensure that the technical infrastructure and processes established for the experiment are sound. Figure 4 shows the processes in this phase. These dry runs must include all components, with rehearsal continuing until all the systems run correctly and reliably. During this phase, the team must never assume that there are individual systems that do not need testing, nor assume a fix to a system is successful until it is tested in an experiment context.

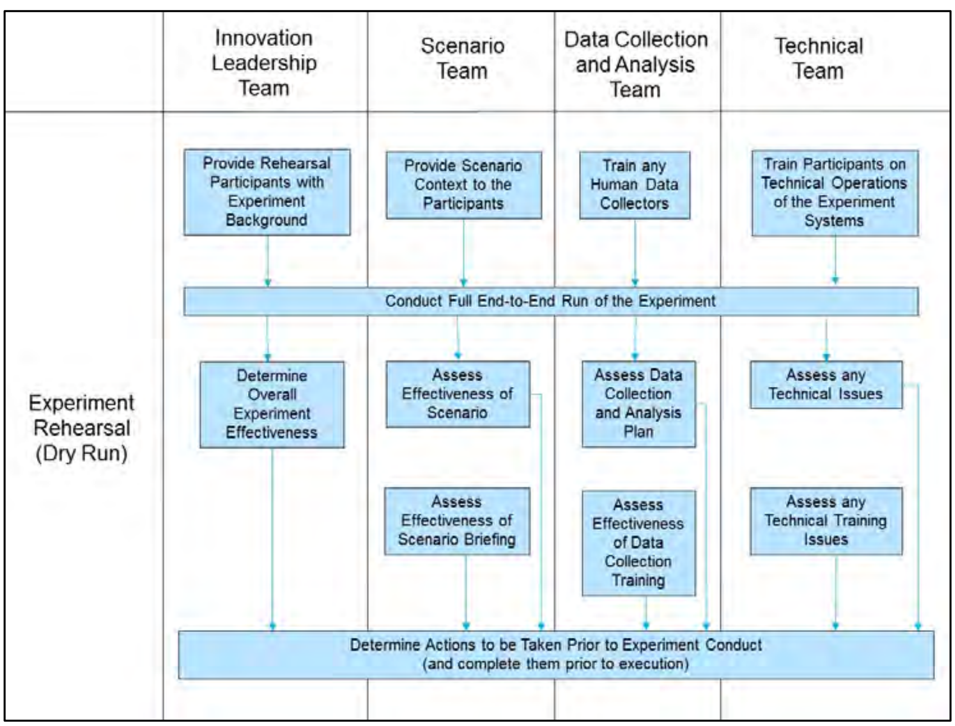

Figure 4: Experiment Rehearsal Processes.

Rehearsals will involve human participants. Ideally, these persons will understand how the systems would be used in real world operations. It is not a good idea to use developers or even experiment designers as the human participants in the dry runs, as they may have preconceived notions of how someone else may use the systems. In addition to human subjects of the experiment, any persons who will be observers for the DCA team should also participate in the rehearsals. Human participants will need to be trained as part of the rehearsal, but this also provides an opportunity to assess the adequacy of the training materials, which will also be needed for the actual experiment. The rehearsal should also be used to verify that the expected data is produced in the proper formats to supports analysis.

Prior to a rehearsal, the scenario team, DCA team, and technical team should conduct an analysis of expected result, so that during the rehearsal, team members can quickly determine if something is going wrong. Otherwise, it will be difficult to assess whether the results generated are reasonable. Thus, the scenario team should have a solid understanding of how the scenario should evolve, the technical team should know how systems will work, and the DCA team will know what data is expected to be generated.

During scheduling, it should be expected that things will go wrong during the dry runs, so there should be enough slack in the schedule to allow time to address issues. Conducting a single rehearsal right before a scheduled experiment will probably not be adequate for ensuring a correct, reliable experiment.

During a dry run, the team should be prepared to stop the test when things go wrong. This is not the time to ignore problems. Instead, an open, collaborative environment among the experiment team and the human participants is important to resolve the issues encountered.

3.7 Experiment Execution Phase: Execution of the Plan

While it is true that the most important part of the experimentation process are the planning and preparation that leads up to the actual execution, it is essential that for execution, all participants are trained properly, and that progress is monitored at runtime. In addition, it is critical that both quantitative and qualitative data is collected during a run and archived at the end of each run. Figure 5 shows processes during the execution phase.

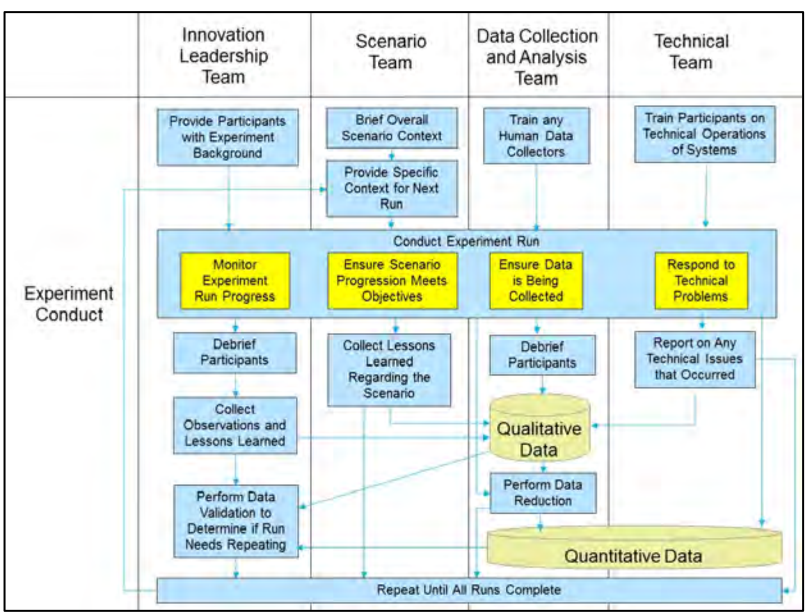

Figure 5: Experiment Conduct Processes.

For successful execution, training must occur for several categories of persons. The experiment operators should be instructed on the experiment purpose, the underlying scenario, and the roles they will be expected to perform (Harvey et al. 2003). The experiment control team must be trained on when and how they will intervene in the experiments. Finally, any observers for the DCA team should receive instruction on what type of information to collect.

As Morcov states, complexity is managed by monitoring controlling, and implementing response strategies (Morcov et al. 2021). During an experiment run, the experiment team must monitor and control the experiment to ensure that it is progressing in a manner that meets experiment objectives. The experiment team should be responsive to problems that occur with the experiment conduct, as some items can be resolved quickly. It is important that the team not wait until after an experiment run to discover that a fixable problem affected the quality of the experiment results, or that the needed data was not being collected successfully.

During each run, the DCA team should perform real-time processing of data to ensure that the data being generated is reasonable. If any problems occur during a run, the DCA team must document the problem and determine if it had a significant impact on results. If so, the DCA team may determine that results and not valid and that the run must be repeated with the problem fixed.

At the completion of a run, several actions must occur. The experiment team should collect feedback from participants while information is fresh in their minds. This can occur in the form of post-run surveys or by gathering all participants together to discuss the run; if both methods are employed, the post-run survey should occur first so that participants are not biased by what they hear from others.

3.8 Post-Experiment Phase: Analysis, Interpretation, and Information Propagation

After the experiment has been executed, data collected from the experiment must be assembled, analyzed, and propagated. This is done largely by the data collection and analysis team. The team must integrate results from multiple sources to gain a strong understanding of experiment results. These sources may include automated data collection, observer notes, audio, video, chat logs, operator feedback, and survey results. To assess results, the DCA team can leverage analysis tools and statistical methods where appropriate. In addition, the DCA team should look for new insights. The analysis must also account for any anomalies caused by experiment glitches.

The DCA team will ultimately generate a final report. Results may include findings and the interpretation of those findings, but the documentation should differentiate between objective and subjective content. The authors of the final report should also describe experiment results in the context of other knowledge, including prior experiments. The DCA team should also consider how analysis results from this experiment could feed into a future experiment.

4 HITL EXPERIMENT APPLICATIONS

This section gives a brief overview of how the above processes have been used, including a description of generalized use cases and the degree of success in managing complexity.

4.1 Examples of Human-in-the-Loop Experiments

HITL experimentation can be used to solve many problems. Below are two illustrative examples.

HITL is very useful early in the Systems Engineering Lifecyle to help define requirements for new systems. For example, user interfaces of proposed software can be improved using HITL experimentation. Developing configurable prototypes for real operators to use can lead to refinement of the software requirements before requests to proposals are sent to industry. During a HITL event, prototype software can easily be modified to meet user preferences. This approach is very cost effective as opposed to having to update software after fielding to users.

Future concepts exploration is also well suited for HITL experimentation. Screening of several potential capabilities can be easily explored during a well planned and executed HITL experiment. There is no need to fully develop new systems to assess how they might operate in a realistic environment. A HITL goes beyond the physics of things as human operators can assess the utility. The human participants often produce new ideas for future systems by having been immersed in the environment.

4.2 Record of Complexity Management

MITRE has evolved the processes described in Section 3 over about 15 years of planning and conducting collaborative simulation-based experiments for a variety of sponsors. Earlier experiments were not nearly as successful as current ones, and the processes documented here are based on lessons learned. Using the methodology described in Section 3, the experiment team has been able to expand the number of experiments supported, going from 2-3 experiments per year to currently 4-6 experiments per year. The conveyance of the process to new personnel has facilitated contributions from many persons.

Understanding and communicating the experiment processes to the persons who sponsor our work has also been effective for expectation management. Sponsors participate in the upfront planning with an understanding of what will be required to meet experiment objectives, and they facilitate the participation of external persons who can support the effort. This external support may include subject matter expertise or provision of technical systems that can be integrated into the experiment infrastructure.

To be clear, though, the process is not easy, and it is not perfect. Collaboration among the experiment sub-teams requires making difficult decisions regarding tradeoffs necessary to make the experiment feasible in the time allotted. There is also some complexity that can still be managed better. For instance, due to the limits of resources for the experiment team, it is difficult to perform full scalability testing of the experiment infrastructure. Often, it is not until all the human subjects show up for the experiment that we can determine how well the infrastructure performs with a full load of operators. In addition, despite our best efforts to perform the necessary planning early in the experiment cycle, there are still times when we must accommodate late requests due to changing circumstances. As there is always room for improvement, the experiment methodology continues to evolve.

5 SUMMARY OF MAIN POINTS

This section emphasizes some key points for the collaborative HITL experiment processes. This section should not be used as a shortcut for understanding the methodology, as many of the details described in Section 3 are critical for experiment success. Having said that, here are points of emphasis for managing the complexity of collaborative simulation-based HITL experiments:

- Upfront, it is critical that the experiment objectives are well defined. All the subsequent experiment planning, conduct, and analysis are derived from the initial understanding of the objectives.
- Key variable relationships must be determined during experiment formulation, so that the experiment team has a clear understanding of independent variables and metrics.
- During experiment planning phases, the scenario, technical, and DCA teams must work together closely. The scenario team must define a scenario that the technical team can implement, and it must be suitable for generating data that the DCA team needs. The technical team must convey to the scenario team what is technically feasible and must ultimately be able to represent the agreed upon scenario. The technical and DCA teams must also collaborate so that as much of the needed data as possible can be automatically generated from the systems in the technical infrastructure.
- The DCA team must create a Design of Experiments that can assess the hypotheses developed during experiment formulation. The DOE must be defined early enough so that the scenario and technical teams know how the scenario will be varied each run.
- During experiment rehearsal and testing, the innovation leadership team, scenario team, and DCA team must ensure that the experiment infrastructure developed by the technical team represents the desired scenario and is sufficient for collecting data to meet the experiment objectives.
- At experiment conduct, all the experiment teams have a role in collecting data for analyzing and documenting results.

ACKNOWLEDGEMENTS

Approved by MITRE for Public Release; Distribution Unlimited. Public Release Case Number 21-2209. The author's affiliation with The MITRE Corporation is provided for identification purposes only and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the authors.

REFERENCES

Alberts, D. S., Hayes, R. E. 2002. *Codes of Best Practice for Experimentation*. DoD Command and Control Research Program Publication. Washington, D.C.

Department of Defense (DoD). 2004. *Defense Acquisition Guidebook*. Washington, D.C.

Girmscheid, G., Brockmann, C. 2008. "The Inherent Complexity of Large Scale Engineering Projects". *Project Perspectives*.

Harvey, A., Buondonno, K., Kopardekar, P., Magyarits, S., Racine, N. 2003. *Best Practices for Human-in-the-Loop Validation Exercises*. EUROCONTROL Experimental Centre, Federal Aviation Administration, and Titan Systems.

Morcov, S., Pintelon, L., Kusters, R. 2021. "A Framework for IT Project Complexity Management". *14th IADIS International Conference Information Systems*.

Popovics, G., Montosori, L. 2016. "An Approach to Determine Simulation Model Complexity". *Procedia CIRP 52*.

The Technical Cooperation Program. 2006. *Guide for Understanding and Implementing Defense Experimentation (GUIDEx)*. Joint Systems Analysis Group, Methods and Approaches for Warfighting Experimentation Action Group 12. Ottowa, Canada.

ADDITIONAL READING

The following references are offered as additional sources that support understanding of experimentation processes.

Alberts, D. S., Hayes, R. E. 2005. *Campaigns of Experimentation: Pathways to Innovations and Transformation*. DoD Command and Control Research Program Publication. Washington, D.C.

Fall, T. 1997. "A Framework for the Simulation Experimentation Process". *Proceedings of the 1997 Winter Simulation Conference*, edited by S. Andratottir, K. J. Healy, D. H. Withers, and B. L. Nelson. Lockheed Martin Western Development Laboratories. Sunnyvale, California.

Murray, W. 2000. *Experimentation in the Period Between the Two World Wars: Lessons for the Twenty-First Century*. Institute for Defense Analysis. Alexandria, Virginia.

Sanchez, S., Wan, H. 2012. "Work Smarter, Not Harder: A Tutorial on Designing and Conducting Simulation Experiments". *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher. Operations Research Department, Naval Postgraduate School. Monterey, California.

Warnke, T., Uhrmacher, A. 2018. "Complex Simulation Experiments Made Easy." *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson. Institute of Computer Science, University of Rostock. Rostock, Germany.

AUTHOR BIOGRAPHIES

DAVID PROCHNOW is a Principal Simulation Engineer at The MITRE Corporation, where he has worked on distributed simulation efforts for over 25 years. His experience has included roles as project lead, technical lead, systems engineer, and hands-on developer, while working in the experimentation, training, test and evaluation, and analysis domains. Over the last 14 years, he has been Technical Lead for one or more collaborative simulation-based experiments per year. Mr. Prochnow holds a BS in Computer Science from the University of Virginia. His email address is prochnow@mitre.org.

ROBERT PORTIGUE is a Principal Multi-Discipline Systems Engineer at The MITRE Corporation where he has worked on HITL experiments for the past 12 years. Prior to that, he was an Operations Research analyst in the US Army, involved in numerous analytical capacities as a Career Army Officer. His experience has included roles as a Project lead, Data collector in both supporting and leading Data Collection and Analyis Teams, Leading Scenario teams and Serving as lead of Innovation Leadership Teams. Mr. Portigue hold a BS in Operations Research from the United States Military Academy and a Master in Systems Engineering from Johns Hopikns University. His email address is rportigue@mitre.org.