10th International Command and Control Research and Technology Symposium
The Future of C^2

Lightweight Synthetic Task Environments for C^2 Research Experimentation

Donald A. Cox, Karol G. Ross, William A. Ross

donald, karol, bross@decisionmaking.com

Klein Associates Inc.
1750 Commerce Center Blvd. N.
Fairborn, OH 45324
937-873-8166 / Phone
937-873-8258 / Fax

March 15, 2005

Abstract

Force Transformation can be seen as a battle against complexity and uncertainty. This is especially true when we consider the less concrete components that are central to effective Command and Control (C^2). Experimentation is the force we deploy in this battle as researchers (Alberts & Hayes, 2002). Our main force is large-scale experimentation and simulation. As with the main force in combat, the energy of experimentation can be squandered if not properly focused. Large-scale experimentation also has a large logistical tail, making it a precious resource in an era of increasing demands and restrictions. As such, we must be sure that it is deployed wisely, to maximum effect. In this paper we describe our lightweight Synthetic Task Environment (STE) concept and how it fits within the experimentation campaign framework. We report on our development and use of the lightweight STE concept. We describe the central construct, cognitive authenticity. We have set forward a process for creating STEs, and a framework for documenting their design. Further experience is needed to learn how to optimize the creation of STEs so that they can keep pace with force transformation efforts in the US Armed Forces.

1. Introduction

Force Transformation can be seen as a battle against complexity and uncertainty. This is especially true when we consider the less concrete components of total systems development such as doctrine, tactics, techniques, and procedures (TTP), and leadership. Experimentation is the force we deploy in this battle as researchers (Alberts & Hayes, 2002). Our main force is large-scale experimentation and simulation. As with the main force in combat, the energy of experimentation can be squandered if not properly focused. Large-scale experimentation also has a large logistical tail, making it a precious resource in an era of increasing demands and restrictions. As such, we must be sure that it is deployed wisely, to maximum effect.

In combat, the concept of light infantry was developed to complement the main force and ensure its proper and effective use. The light infantry moves quickly to find and fix the enemy so that the main force can be brought to bear. In this paper we present the beginnings of the development of Synthetic Task Environments (STEs) as the light infantry for the battle of Force Transformation. Synthetic Task Environments are minimalist simulations that reproduce only those aspects of the world needed for the cognitive performance the researchers are investigating. Providing insight and understanding and thus sharpening the conversation about the possibilities and constraints in force transformation is the role we see for STEs.

Our experience with large scale experimentation for doctrine and concept development in Future Force Unit of Action (UA) and Unit of Employment (UE) C^2 has been both very exciting and very frustrating. It is exciting for us because we can see our theories and predictions enacted in both expected and unexpected forms. We were able to gather a great quantity and variety of data. However, it was frustrating for a number of reasons. Perhaps foremost among these was a certain “look but don’t touch” vibe from some of the experiment controllers which is antithetical to our approach as naturalistic researchers. While this is understandable given the pressure on them to produce a successful event, we see it as interfering with the ultimate success of the ostensible purpose of the experiment. The seemingly inevitable technical and organizational glitches also distracted from the experience and diminished our confidence in the data. And finally, the fact that these events are few and far between led to many researchers trying to collect many data, sometimes at odds with each other. We saw the STE concept as promising a framework that would allow us to preserve what we found exciting, while ameliorating much of what was frustrating in a systematic program.

In this paper we lay the groundwork for STEs to play the light infantry role in force transformation. In order to successfully deploy STEs as a lightweight experimentation concept within the force transformation effort, we must address the following issues:

- What is an STE?
- How do we create STEs?
- How can we create STEs quickly and efficiently?
- How can we have confidence in the results of STE-based experiments?

We will focus on the first two issues, with suggestions about how to address the last two issues.

Force transformation is a highly complex process that needs principled experimentation in order to succeed. There are more issues and interactions that need to be explored than can be addressed using traditional exercises and simulations. We believe that lightweight STEs can be the experimental vehicle to address these issues and investigate the interactions. Lightweight STEs are cognitively authentic minimalist simulations reproducing the relevant feature of the setting being investigated. In order for lightweight STEs to successfully contribute to force transformation in this role, we must have an open, repeatable, validated process for creating them that can be followed by any researcher.

Section 2 discusses in more detail the challenges in force transformation and experimentation. The published research on STEs is reviewed in Section 3. Section 4 describes our process for creating STEs. A framework for the critical step in STE creation, abstract design, is detailed in Section 5. Section 6 sketches a case study in creating an initial abstract design for an STE to explore transformation in Army Aviation Command and Control (A2C2). The next steps to fully fleshing out the lightweight STE concept are covered in Section 7.

2. Experimentation for Force Transformation

Force transformation aims at revolution. It seeks to change not one thing but many things at once. There is a hypothesized future in which new systems, not yet developed or even fully specified, will be used by soldiers in ways very different from those of today. The problem of designing for force transformation reflects the envisioned world problem (Roesler, Feil, Woods, Puskeiler, & Tinapple, 2001). We cannot predict all the factors in an envisioned world and how they will affect actual performance. While we don't know for certain what will happen when the future addressed by the envisioned world becomes the present, we can still produce useful guidance, insight, and framing of issues and trade-offs. It is impossible to predict exactly what will happen because of all the different factors that go into making a system such as Future Combat System or a program such as Transformation a success or failure. Although we cannot predict with certainty what will happen, we can do much that will contribute to the successful management of uncertainty and risk in the unfolding development of these systems and concepts.

Experimentation has rightly been identified as a key activity in orchestrating force transformation (Alberts & Hayes, 2002). There are many players who must be brought together to achieve a harmonious result. Developers of doctrine, organization, training, leadership, material, and soldiers must work together to follow a single goal in a dynamic environment. The choices of one developer are connected with the work of the rest. As we have learned in the development of systems, a waterfall model where one developer makes choices without regard to the consequences for the others will lead to large-scale inefficiencies and diminished effectiveness of the resulting product. Experiments provide the touchstone of common experience that coordinates all the developers.

Experimentation is often taken to mean large-scale exercises or simulations. These events that involve multiple systems and multiple participants are complex. They are complex to organize and fund, so they can only occur infrequently. Because they are infrequent, they are valuable. They often try to address a large range of issues because everyone wants a piece of the action. This leads to a variation of the "tragedy of the commons" (Hardin, 1968) where the experiments have so much going on it is hard to tell what has happened—there are many potential confounds. Nonetheless, well-executed large-scale experiments are necessary to see the dynamics that appear only when the system-of-systems is operating in all its complex glory. The DoD and Uniformed Services recognize this and have made extensive efforts to model and simulate the envisioned world of the transformed force. Unfortunately, these large-scale activities often produce ungainly and unfocused amounts of data, as well as new issues that are difficult to address.

The other usual connotation of experimentation is in some ways the exact opposite. It is a carefully controlled situation that seeks to describe the causitive, or at least correlative relationship between a limited number of carefully defined independent and dependent variables. These experiments can produce deep insight and elegant results. But they can also be no less resource intensive than large-scale exercises. And it can be a challenge to translate their findings into practical knowledge of how to proceed in systems design. We value both of these approaches, but see the need for a middle way. We are developing the lightweight STE concept as that middle way.

If these experiments based on large-scale exercises or simulations are the main force of experimentation, STEs can play the role of light infantry, conducting reconnaissance surrounding the experimental effort. This use of STEs allows experimenters to see where to focus their experimental power and to examine what has happened afterwards. Synthetic Task Environment experiments can be run in advance of a large-scale experiment to explore the issues upon which the main event intends to focus. Results from the STE experience can be used to guide the main experiment to thoroughly address the issues, for instance by verifying the conditions needed to evoke the cognition or other phenomena the experimenters wish to study. After the experiment, STEs can be used to more closely explore issues that came up during the experiment. All good experiments suggest areas for follow-on study. Synthetic Task Environments can be used to begin the exploration of those areas, for instance to more closely identify the parameters that affect the phenomenon observed in the experiment.

Conducting research to evaluate the effects of the transformation process on the warfighting system-of-systems poses a significant challenge and requires a relatively low cost, quick turn-around approach to system prototyping. We believe the STE concept is a high value asset in the race to transform military organizations, systems, and soldiers from linear to network-centric, system-of-systems warfighting teams. The forces of transformation and spiral development are relentless and demand fresh approaches for researching untested concepts and methods.

3. STE Research to Date

Interest and activity in the development and use of STEs arises from two different perspectives. Researchers into the nature of cognition and cognitive performance are interested in STEs as a middle ground between highly controlled and contrived laboratory situations and rich and complex real-world situations. In this perspective STEs promise to be rich enough to engage the complex cognitive phenomena and expertise, yet be simple enough to be manageable for traditional experimental analysis. Investigators' primary goal in using STEs is to better understand the nature of cognition. This perspective has been the main area of activity in the STE literature.

The other perspective that finds STEs interesting is applied research. The primary goal in applied research with STEs is understanding the consequences of conceptual and technical innovation. The appeal of STEs from this perspective is as an efficient experimental situation that can evoke realistic cognition and performance without the complexity and overhead of large-scale simulation or exercises. The use of STEs in this context also appeals to our stance as researchers in naturalistic decision making and related cognitive phenomena. We believe that performance and expertise are connected to the environment of performance. While we can gain some understanding of human cognition from a random sample of people, to understand expert performance we must study performers in the real world. We are more focused on seeing real performance than establishing control.

Force transformation is a wicked problem (Rittel & Webber, 1973). There is no one right answer. We are not faced with the question of what is the best technology to support a particular doctrine, or what is the best doctrine for employing a particular technology. The best case scenario is that doctrine and technology are co-evolving with advances and failures in one impacting the other. We need to explore the space of possibility and consequence. Even if theories of cognition were much more advanced, they would still only be a piece of the puzzle of successful force transformation. The applied perspective, while it may not deliver the truth, is our best hope for providing guidance and having open conversations about the opportunities and risks of force transformation.

The research and applied perspectives are not mutually exclusive. They share many concerns. Foremost is that results are valid and reliable. But in speaking to different audiences for different purposes, they must have different priorities and presentations.

Researchers have found STEs to be a valuable method for studying cognitive and behavioral task performance in a variety of domains such as aviation (Flach et al., 2003), Uninhabited Air Vehicle (UAV) operations (Cooke & Shope, 2004; Gugerty, DeBoom, Walker, & Burns, 1999; Martin, Lyon, & Schreiber, 1998), sonar operation (Gray, 2002), and AWACS Command and Control team environments (Hess, MacMillan, Elliott, & Schiflett, 1999). There are many benefits associated with using STEs to explore specific research questions. Foremost, STEs strike a balance between the qualities of field research and laboratory experiments. Properly designed STEs retain the fidelity of the real-world task, and the experimental control and generalizability of laboratory research. However, establishing the validity of STEs has been challenging, and the issue of fidelity is a constant topic of discussion in the literature.

3.1 How Do You Create an STE?

All the STEs are created using a similar high-level process. The real-world performance in a related area of interest is studied using Cognitive Task Analysis (CTA). This analysis is used to select portions of the real world and the task that are connected to the cognitive phenomena of interest to the researcher. The researcher constructs the STE and engages subject-matter experts (SMEs) to assess whether it works. However, to create lightweight STEs we need a more detailed description of the process that does not depend on the cultural lore of human factors researchers. As described later in this paper, we put forth a process that also follows that structure, but provides a greater level of detail and direction than previously described.

To date, Cooke and Shope (2004) have provided the most detailed description of procedures used to create an STE. First, the researchers gathered information to constrain the design, which included gaining an understanding of the research domain, identifying research goals, and identifying additional constraints. Next, based on their domain knowledge and research objectives, the researchers abstracted features of the task and included them in the STE. Task elements related to team cognition, such as knowledge and information sharing, planning, and dynamic replanning were preserved in the STE design. Cooke and Shope could not state the procedure they used to abstract these features. They described it as seeming like "intuition or art, rather than the application of well-specified rules" (p. 270). They explain that their experience and domain knowledge were probably the basis for these decisions. The researchers then built a prototype, including paper mock-ups and functional specifications. The final stage centered on implementing the design and iteratively redesigning the STE based on reviews.

The Martin et al. (1998) procedure included conducting a CTA and then abstracting (they used the term decoupling) the key aspects of the task from the environment. They began by conducting interviews with domain experts. From these interviews the researchers identified the goals, cognitive demands, and resources required to perform the task. Then, the high-workload, high-skill parts of the task were identified and filtered from the task environment. The researchers did not provide guidelines as to how to filter these key task elements from the environment. The synthetic task was tested, and design problems were resolved through an iterative design cycle that included determining the validity of the STE prior to implementation. Informally, validity was established by having expert Predator operators review the STE during demonstrations and interviews. Martin et al. (1998) assessed validity through experiments with experts and novices. The reasoning behind this method of empirical validation was that if the researchers were successful in preserving the key cognitive and behavioral elements of the task, experts would have higher performance than the novices.

3.2 Ensuring Quality Results

Synthetic Task Environments have the same requirements for validity and reliability as any other experimental setting used in scientific inquiry. In an applied setting, we are particularly concerned with defining a process that does a good job ensuring these properties are in effect so that time and effort need not be spent in repeatedly verifying the STEs generated. Synthetic Task Environments face particular challenges in establishing validity because they aim to occupy the middle ground in a spectrum that is not well defined. As described in the previous section, the creation of STEs has relied on considerable judgment on the part of the researcher. As such, we need a way to check the intuitions of the STE designer.

There are a number of ways that STE researchers have gone about establishing validity. Review of the STE design and implementation by subject-matter experts is a popular method for verifying face validity. This procedure is used in Gray (2002) and Cooke and Shope (2004). We continue to recommend this practice in our STE development process. Additional guidance on how to make this review efficient and effective is needed.

An additional technique for establishing validity is to see that experts perform better experimentally in the STE than novices. This procedure is used by Martin et al. (1998). If the experts who participate in this experiment are the same ones who are involved in defining the STE, this result is not particularly powerful, as it appears circular. This is a useful result nonetheless, as experts' descriptions of how they carry out their work are often at odds with observations of how they actually perform. However, validation with a random sample of SMEs would provide a more solid basis.

The judgment of what is relevant and necessary to preserve or discard is also very theory-laden. Without solid theories of cognitive performance, different researchers make different choices. For instance, according to Gugerty (2004), "three essential features of synthetic tasks are that they represent the key cognitive and perceptual demands of a complex task, are easily reconfigurable for research and training purposes, and can be used by non-expert operators" (p. 240). Cooke and Shope (2004) agree. Since their laboratory relies on college students as participants, "the necessary STE skills and knowledge should be learnable with little or no background knowledge and without extensive training and practice" (p. 269).

From our perspective, once we move into the realm of the ubiquitous college-student-as-research-subject, we have left the field and re-entered the laboratory entirely. We believe the STE serves a function between the laboratory and fieldwork that is not filled unless actual intended systems users are STE participants. Similarly, we "draw the line" at including microworlds in our definition of STEs, at least as they have historically been developed to focus on complexity in decision making without regard to real-world contexts (Brehmer, 2004).

The validity of an STE is highly dependent upon the procedures used to develop it. In particular, identifying the elements related to the behavioral and cognitive aspects of the task and preserving their relationship in the STE is both the most difficult and the most important part of the procedure for ensuring fidelity to the original task. Martin et al. (1998) acknowledged that, "Decoupling is crucial. A synthetic task will only be successful if it can be excised from the context of the larger task in which it is embedded without markedly changing the way people perform the task" (p. 126). Few researchers have described the processes they used to create the STE, and the descriptions are often provided at a high-level. A different set of researchers may have difficulties replicating the procedures and creating an STE. Recently, Cooke and Shope (2004) provided a detailed description of their 4-step STE creation process; however, they focus on the technical aspects of creating their STE hardware/software system, and the filtering process (Cooke and Shope use the term abstraction of features) still remains vague. To increase validity and the ability to inspect and verify it, we propose a more detailed design process and documentation framework for this key translation step.

As Cooke and Shope (2004) observe, there is a connection between fidelity and validity in the construction and use of STEs. Fidelity is a quality of the translation from real-world performance to the STE, and validity is a quality of the translation from the STE to the world of performance. Ensuring fidelity will help establish validity, especially in applied contexts. In our STEs we are concerned with observing cognitive performance and so concern ourselves with cognitive authenticity (Ross & Pierce, 2000). By this term we mean the preservation of critical perceptual cues and contextual factors. Critical cues and factors are those that cause a participant to feel as if judgments and decisions are real and important. Furthermore, these cues and factors enable the naturalistic cognitive and perceptual processes to be stimulated and engaged in the environment. The achievement of such authenticity is not necessarily dependent on an exact representation of all dynamic features in the natural environment. We refer to authenticity and not fidelity because authenticity imparts a connotation about the nature of the experience during performance. A dynamic quality is created by cues and factors in the environment that is stimulating the cognitive processes. Fidelity in experimental settings is usually taken to refer to the concept of physical fidelity that drives most simulation development, and the connotation is that a match between physical characteristics of the real world and the simulated world are needed. Our experience is that a high degree of physical fidelity is not needed to obtain cognitive authenticity. Physical fidelity requires a heavy investment of resources that can be better spent financing a campaign of lightweight STE experimentation that will provide equal or better results for informing and managing force transformation.

4. Synthetic Task Environment Development Process

In order for lightweight STEs to successfully contribute to the exploration and definition of force transformation, we must have an open, repeatable, validated process for creating these STEs that can be followed by any researcher. In this section we describe our variation on the standard development process. We have inserted specific guidance about how to carry out these steps when creating lightweight STEs.

In our conception, there are broadly four phases to the development and use of STEs. They are as follows:

1. Data collection – gather the resources necessary to define the research question and task and environment in which the question is to be investigated.
2. Analysis – systematically structure the data gathered to capture and relate the critical cognition for the question being investigated. We refer to this structure as an abstract STE. We put forward an STE design framework to document the abstract STE and the critical design decisions made in constructing it.
3. Construction – build one or more concrete environments in which the critical cognition is necessary and sufficient for task completion.
4. Experimentation – use the concrete STEs for experimentation to advance towards an answer to the research question.

We see the contribution of this paper not in identifying these generic product development steps but in putting forward a level of detail not previously published. We will now consider each of those phases in more detail, especially the analysis phase and the use of our STE framework.

4.1 Phase One – Question Definition and Data Collection

Data collection is the key to abstracting the key decision tasks from the envisioned context. An STE that does not include the meaningful aspects of the performance context cannot support valid measures of performance, usability, or acceptability. When the performance context demonstrated within the STE is representative, the results of STE experiments minimize highly ambiguous or uncertain situation aspects of the future situation. The application of CTA at the front-end of the STE process enables the researcher or developer to focus on the most critical aspects of complex performance as actually carried out in the real world.

Our data collection phase focuses on collecting data about the critical cognition in the performance we seek to study in the STE. We conceive of this phase consisting of the following activities:

1. Identify the research question. Our desire to limit the complexity of the STE will mean that we must have a focused research question. We will use Macrocognition (Klein et al., 2003) as a guide for helping us select and focus our questions.

2. Investigate current or analogous performance. STE designers must investigate the phenomena the experimenters wish to explore in order to locate the parameters for the STE design. This data will be analyzed in the next phase to create the core of the STE design. The STE will only incorporate elements that are captured in the data collection. For Klein Associates, part of this process would be conducting or reviewing CTA interviews with relevant SMEs to identify goals, cues, cognitive demands, and resources relevant to the situation under study. We recommend data collection be carried out with an eye towards the analysis that will be conducted on the data.

The steps within this phase may have to follow an iterative or spiral development, as what you learn can change your question. In our experience, when you are new to a domain or area of investigation, what you initially believe is a small question upon further investigation can grow beyond your ability to address it comprehensively. This process forces the STE to focus on a more precise question than the one you started with.

4.2 Phase Two – Analysis

The data collection phase gives us a point of view and a picture of the world. The goal of the analysis phase is to identify both the critical cognition that occurs in accomplishing the task under study and the features of the situation necessary to enable that cognition. This is the critical phase in STE development, as we need to progress in a way that we can be confident we are maintaining cognitive authenticity and that others can audit our work to check our claims. This is done through an activity we call abstract design.

3. Select which features of the reference situation to preserve and which to simplify. This is the where the most consequential decisions are made in STE creation. The design challenge for a lightweight STE is to have an environment rich enough to give rise to the phenomena to be studied, but simple enough to be implemented with the resources available. These choices should be justified on the basis of the experimenter's models and theories of performance. We typically refer to this challenge as maintaining cognitive authenticity—ensuring that what we produce poses the same cognitive challenges as the real-world situation, and engages the same expertise. For example, if we were creating an STE to study decision making, then we would have created Decision Requirements Tables (DRTs) which capture the critical decisions and the cues, factors, and strategies used by the decision maker. To create the STE, we need to look at the DRTs and re-represent what is captured there in a format that can be used more directly in design. We have created a framework for creating and capturing the abstract design of the STEs and this framework is described in a later section of this paper.

4. Next we vet the abstract design. This is early peer review and there are two communities of peers for us to address. The first is researchers, who inspect the abstract design for theoretical soundness and experimental power. The second community is SMEs from the domain or domains in which the STE is set, who provide a reality check and face validity. The minimum review is by one SME in each category. Ideally, we would like to have three SMEs from each category look at the design at this point.

The output of the analysis phase is an abstract STE design. In software development terms you can think of it as the requirements or architectural specification for the STE. Multiple concrete STEs can be produced from a single abstract STE design. The abstract STE captures the information, affordances, and dynamics that are necessary for performance to occur.

4.3 Phase Three – Construction

This is the phase where we make the abstract STE concrete, put flesh on the bones, and turn it into something people can actually use. There are four stages in the construction phase.

5. Initial design of the STE. We must choose how to instantiate the abstract STE into a concrete form that will provide the qualities that Gray (2002) mentions: face validity and realism, participant engagement, and experimental ease of use. The primary decision in this stage is the determination of the nature of the STE—is it one to explore individual cognition, teamwork, or new system design and capabilities? This decision will control what is encoded within the STE and what the participants must bring. If the STE is going to be based on a computer simulation, the output of this stage is a functional specification for that simulation.

6. Inspection of STE. Various SMEs must review the STE design for face validity and realism, participant engagement, and experimental ease of use. This would include SMEs from the domain or domains participating in the STE as well as experts in the underlying theory and experimental design. At this stage we recommend that three domain SMEs review the design and that at least one SME in STEs and experimentation inspect the design. If the comments from the SMEs indicate problems in the design, it is advisable to do further design work before continuing, possibly conducting further data collection and analysis as indicated. Once the SME review feedback reflects sufficient confidence in the STE design, actual STE construction can begin.

7. Complete construction of STE. The STE must include a description of the task, an interactive environment, and data collection facilities. The STEs reported in the literature are heavyweight set-ups based on computer simulations. We believe that lightweight STEs can be built either without simulation support, or with minimal support. If sensory-motor performance is critical to the STE design, then simulation will be needed to provide adequate responsiveness. However, there is a large class of problems which can be adequately and realistically dealt with by Wizard-of-Oz techniques (Maulsby, Greenberg, & Mander, 1993) where one or more human confederates of the experimenter provide the dynamics of the environment. These confederates would be provided with the necessary scripts and guidance so that STE performance would be reliable.

8. Pilot test of STE and iteration of STE design. As with any experimental design, after development and construction, the first activity is to pilot test. A small sample of participants should try out the STE and examine the results. If the results are far from those expected, the earlier phases and activities should be revisited to ensure they are well-justified and reasoned. At this point it should be verified that experimenters can get good data, and that the STE functions as expected.

4.4 Phase Four – Experimentation

As with every other form of research, when we conduct our STE research, we must select participants with due care and be mindful of interpreting results to not exceed the strength of the experiment. However, the lightweight nature of STEs should allow the experimenters to run a series of events, allowing for reproduction of the original results and investigation of systematic variations that will increase confidence in the results. In some ways, getting positive results from the first experiment with a new STE is the most troublesome result. The inclination is to accept these results as vindication of the STE design and experimental program. However, the best response is two-fold: first to re-examine the design and implementation of the STE to ensure that it is not a leading confound in itself, and second to engage in further experimentation to replicate or refine the results. If the results of the STE do not match expectations, the response is much the same: re-examine the STE design, implementation, and experimental assumptions for flaws and also conduct further experimentation with the STE to understand the results in a larger context.

4.5 Phase Five – Campaign Continued

Having created the STE and run an experiment in it, it is time to use the results and leverage work done to date. The next step will depend on considerations beyond the particular research effort. Ideally, researchers would like to rerun the experiment to establish reliability through replication. Systematic variation in the experimental conditions will also increase the insight gained from experimenting with the STE. Additional concrete designs may be generated from the same abstract design to explore related cognitive phenomena. For instance, if our first concrete design is to study coordination between the UA and the UE, we may create a second concrete design to investigate decision making within the UE—how

UAVs are tasked and how competing and changing requests and priorities play out at the UE level. Having established our beachhead STE, we seek to expand along two dimensions – the domain and the cognitive phenomena.

5. Framework for Abstract STE Design

Force transformation is a highly complex process that needs principled experimentation in order to succeed. There are more issues and interactions that need to be explored than can be addressed using traditional exercises and simulations. There are also many stakeholders in the process. Critical to the success of lightweight STEs in force transformation experimentation is that there is a place and a process for involving all the stakeholders in the creation of the STE so they have confidence in the results of the experimentation. The abstract design step and the framework for documenting the abstract design presented in this section are the process and place for bringing everyone together and coming to agreement.

Recall that in the analysis step of the STE development process, we analyze the cognitive performance. This breaks down performance into its constituent parts. In order to design an environment, we need a synthetic activity that brings the parts back together again. The creation of the abstract design is this synthetic activity, bringing the elements torn apart in analysis back together into a whole that captures the nature of the work. We have developed a framework for capturing the abstract design that aims to put the pieces back together again. It serves as a way of capturing the key design decisions by creating the STE in a way that presents them directly so that others may review these decisions.

The crux of STE design is the selection of features to be retained and those to be left out from the actual situation of performance. Abstract STE design begins with going back to the data collected and identifying all the elements that are relevant to accomplishing the task to be studied. These elements are all the factors that contribute to the performance under study. The initial pass through the data will yield a large collection of elements. At this point it is tempting to stop and brainstorm what it is about each element that is important, annotate the elements with that list and declare the abstract design complete. But before elaborating on an element, it is important to consider what aspects of the element are relevant to the performance being studied. To facilitate development, we have devised an STE design framework. The framework raises issues that must be dealt with in the design and construction of an STE. We find it is best to deal with those issues at the abstract design stage rather than leaving them until later in the process where they can become confused with details, such as the nature of the situation or narrative. The goal is to develop the collection into a coherent whole that suggests the nature of the performance being studied. To assist in this, our framework uses three perspectives on performance.

The first perspective is an extension of an existing paradigm in Cognitive System Engineering that says performance arises out of the interaction of people, their work, and the technology or tools available to them. Those three categories are adequate when considering activities such as flying or nuclear reactor operation where the performer is dealing with a physical system. To get a fuller and more realistic picture of performance in a domain such as military operations it is necessary to add a fourth category, namely environment, to account for the fact that performance happens in widely varying conditions such as the terrain, the local population, and the enemy. We use the people-work-technology-environment perspective to organize our STE elements. We use this perspective to classify elements. It also enables us to discover new elements we've previously overlooked by asking of each of the four categories, "What more is there in this category that is relevant to performance? Are there other aspects of the environment, other technologies used, other people contributing to the performance?"

The second perspective on performance is that it is the interplay between goals, capabilities, and constraints—how can I achieve my goals with the capabilities at hand given the constraints on my actions? For each element we can ask what goals, if any, does it have. What constraints does it impose?

What capabilities does it provide? Not every element will have an answer to all those questions. For instance, if we have an element representing weather in our STE it is not useful to identify a goal for weather. The goal of these perspectives is not to provide a spreadsheet that will allow the STE designers to calculate what they should do. Rather, they are cognitive prostheses that exist to help the STE designers capture, extend, and reflect upon their thinking. In this perspective we also distinguish a special sort of capability, namely who or what the element communicates with.

The final perspective is that of STE implementation. In this perspective we categorize elements into actors, resources, environmental features and concepts. Actors are elements that behave independently of the rest of the STE. Actors that can be "played" by human participants are identified as an "Actor: Role". An example of a non-human actor would be the weather. Resources are tools or technologies that the actors use to discover and change the state of the STE. Environmental features are inert elements such as terrain. Concepts are elements that are not implemented directly in the STE. In the UAV example the UA is modeled as a conceptual element—there is no UA in the STE, though the notion of the UA is important to understanding the performance in the STE, so it is captured as a conceptual element in the abstract design.

What constitutes an element will depend to a degree on the situation and research question. In the UAV example we identify a number of elements involved in planning, allocating and controlling UAVs, as well as distinctions among various UAV types. This focus is the result of the research question and how we believe coordination happens. If we were doing an STE to explore how a UA commander can optimally arrange supporting fires, it might be adequate to say simply that UAVs are a single element and they are under the control of the commander, as all that might be relevant is whether they can provide laser designation or battle damage assessment.

In the first pass through abstract design, we identify elements as they appear in the data collection. Our second pass formalizes the element definitions using the form shown in Figure 1.

ID	
Name	
Element type	
Performance	
Goals	
Constraints	
Capabilities	
Communication	

Figure 1. Element analysis form for STE development.

Our framework forms a basis for STE design and development by analyzing the elements in terms of what causes complexity for the performer and how their salient features for the performer may be captured and described. Our framework consists of filling out, as appropriate, the form in Figure 1 for each of the elements discovered in the data collection phase. We manage the level of complexity and reduce it in our STE design by putting more or less into the fields of the form or choosing a less interactive or complex element type for that element. An example of the latter would be to model the weather as an environmental feature rather than an actor, thereby simplifying both the implementation of it in the STE and its contribution to performance in the eventual STE.

The framework does not prescribe an order in which to complete the fields of the form. We expect that it would be an iterative process—where work on one element or field will influence others as the significance of the changes or elaborations are traced through the whole design. To fill out every field in every form would be inefficient and counterproductive, so the analyst is encouraged to pick those elements and aspects of most interest (as suggested by the research question). Then the analysts only fills in as much of the rest of the table as is needed to support the particular performance at which the STE is aimed. We all have the experience of operating in a resource constrained research situation, so we are in sympathy with this approach. In such a situation, we believe in the continued value of the matrices as a way to provide a check for the analyst and reviewers to see what was done and to suggest other things that might be done as well. Having the matrices provides a place to reflect on how to add or remove complexity if the initial STE design in unworkable.

6. UAV STE Development Example

We now present an example of how we have used the STE design framework. We first sketch out how the process unfolded, and then present in more detail how we have iteratively refined the abstract design. This STE is a work-in-progress, and we document our evolution of the design to date. Our goal in presenting this story of development is to encourage others to try STE development. Developing STEs is not rocket science, but it is patient and careful work.

Our development effort started with discussions with the U.S. Army Battle Command Battle Lab at Leavenworth. They suggested that airspace management in Future Force was a fruitful area for investigation. Future Force envisions an explosion in the number of UAVs operating in the UA and UE airspace. Managing their activity in conjunction with many kinds of munitions as well as manned aircraft is a challenging issue.

Our next step was initial interviews with experienced airspace managers at Fort Rucker. Based on an informal analysis of these interviews, as well as background reading on the issues in managing UAV use of airspace, we decided to further refine our focus for the STE.

Our research question became, “How are the UA and UE going to coordinate their UAVs in complex urban terrain?” This example illustrates the use of our STE Design Framework as part of the analysis phase of the development process. Our first step in creating the abstract design was to draw a diagram of the central players in the question. We created three elements, as shown in Figure 2. Although we do not have ways of representing the abstract design graphically yet, these illustrations were created because they provide a concise way of seeing what is going on in the design and picturing its evolution.

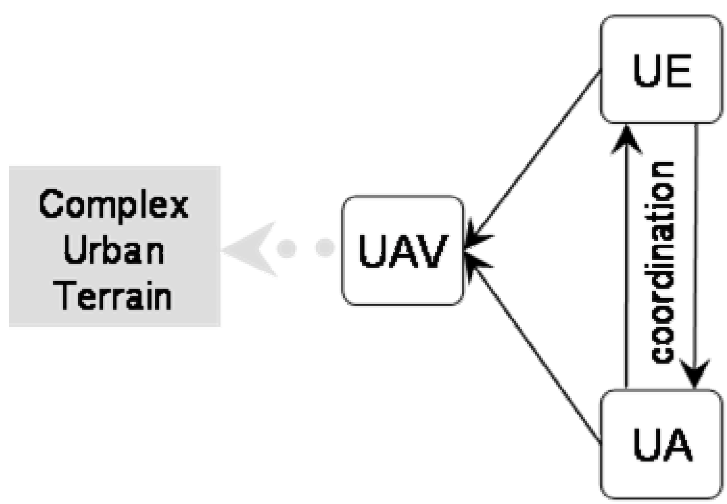

Figure 2. Diagram of elements in research question.

Our next step was to formalize those elements. A pair of the elements is shown below:

ID:	**U2C-1**
Name:	UAV
Element type	Concept
Performance	Technology
Goals	
Constraints	Location Status Mission effectiveness
Capabilities	Sensor package Target designator Weapons package Loiter time Maximum airspeed Minimum airspeed Frangibility Aerial control / manned-unmanned teaming (MUM) Location identifier IFF
Communication	

ID:	**U2C-2**
Name:	UA
Element type	Concept
Performance	
Goals	
Constraints	
Capabilities	Controls own UAVs
Communication	Coordinate with UE

We designated all of these as concept elements because we knew that we were going to have to get more specific. Nonetheless, they formed a good starting point for capturing the essence of the STE, and they were a nice way to enter the model.

Note that the element definitions do not require all fields to be filled out immediately or ever. They exist as cognitive prostheses—to aid our thinking and to communicate it to others. Our approach is that if the formalism is getting in the way of our thinking, then the formalism should be relaxed until we're ready to go back, which may be never.

For our next step we elaborated on the UA and UE. Drawing on data collected on airspace management as well as from a related research effort on tactical thinking and background knowledge of military organization, we generated Figure 3.

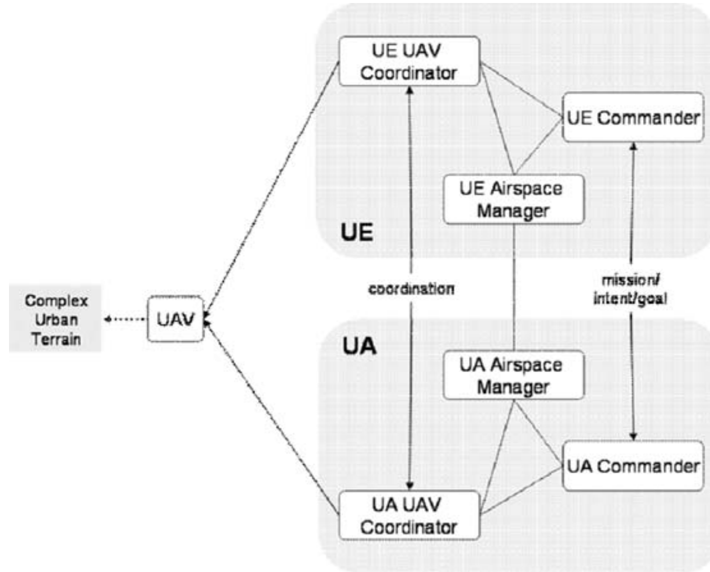

Figure 3. Unpacking the conceptual UE and UA elements.

We formalized those elements in the fashion shown below:

ID:	**U2C-4**
Name:	UE Commander
Element type	Actor: Role
Performance	People
Goals	Win the fight
Constraints	Higher intent
Capabilities	Develop intent, concept of operations Set airspace priorities Accept or reject risks of airspace conflict
Communication	Higher (JTF Commander) UA Commander UE Airspace Manager UE UAV Coordinator

ID:	**U2C-5**
Name:	UE Airspace Manager
Element type	Actor: Role
Performance	People
Goals	Arrange airspace users according to priority to get maximum effectiveness
Constraints	Approval from airspace control authority UE Commander airspace priorities
Capabilities	Mesh airspace users according to airspace priorities minimizing risk
Communication	Airspace control authority UA Airspace Manager UE Commander UE UAV Coordinator UE users of airspace

In our current state of development we have moved on to the other side of the diagram and work on the UAV concept, producing Figure 4 below.

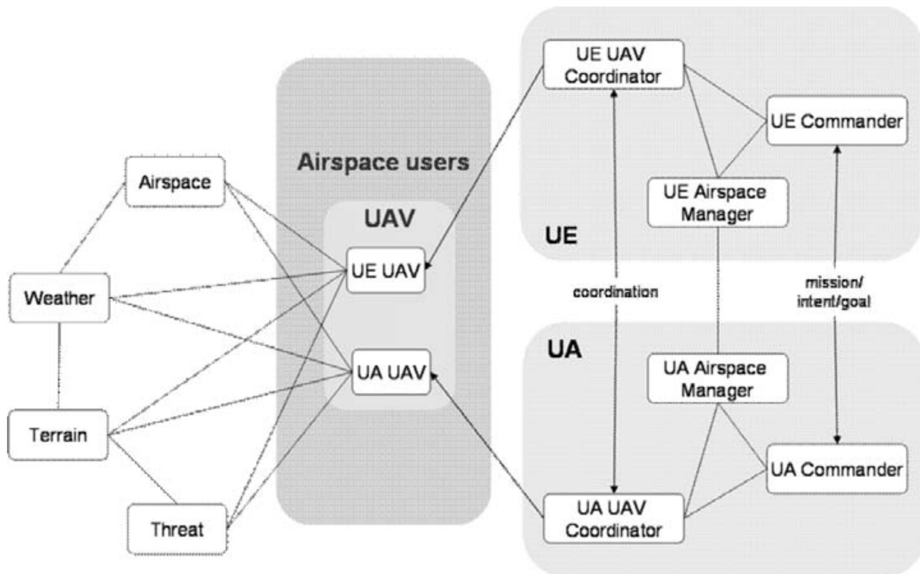

Figure 4. UAV and related elements further detailed.

We have had an initial review of the abstract design at this stage of development. Our next step is to further develop the abstract design to capture more of the relevant elements. For instance, we expect to provide more details about the capabilities of the various UAV classes. We would also like to perform additional interviews with expert UAV mission planners to understand how they manage the constraints under which they operate. Once we have an abstract design we believe captures all the relevant elements, we will conduct systematic walkthroughs with SMEs to validate the design. We will then continue with the rest of the STE development process.

7. Further Development of Lightweight STEs for Experimentation

Our lightweight STE concept is early in its development. We have seen a great need and demand for this sort of experimental tool. In order to deliver a useful tool to the research and practitioner community we have identified a number of directions that need to be taken in order to fully develop the lightweight STE concept. They are to validate the STE process and sample STEs against real-world performance, explore the limits of STE utility, further refine the STE development process, and refine STE creation so that it can be used in a timely fashion by any researcher working on force transformation.

A critical area for future research on STEs is to validate the results against real-world performance. Synthetic Task Environments validation has typically been done by expert review (Cooke & Shope, 2004) or by observing that expert performance in the STE is better than novice performance (Martin et al., 1998). This is not a straightforward task because of the complexity of real-world performance, the number of confounds, and the practical constraints of availability of real-world performers at various skill levels, especially experts. It is precisely because it is so hard that this is what we must push for as a research community. We cannot do it without the support of the larger community of funders and practitioners. As Cooke and Shope (2004) note, the ultimate verification of external validity is comparison of results to those obtained in the real world of performance. More efforts such as theirs need to be undertaken to obtain this verification. We need a larger experimental campaign to validate an STE development process so that the community of researchers and practitioners can have a reasonable level of comfort with the results of STE experiments when a particular process has been followed without having to go through the expense of validating every STE. We should still validate a sample of STEs to make sure that the process is repeatable and sufficient to produce valid STEs. By developing a robust process and a large community of practice, we seek to balance the need for speed in creating STEs with an assurance of quality results.

With envisioned world STEs, the option to compare against real-world performance is not as readily available. Theoretically, one can wait until the real system is available and expertise in it has been developed, and then check the results of the STE experimentation against the real world. This should be done, but it will not address the more immediate and real concern of how much stock to put in the STE results right now, since the reason you're doing the STE experiment in the first place is to make a decision or provide guidance to the development of the system. Waiting to produce a result renders the experiment worse than useless. We need to provide a warrant of the validity of the results before we compare with the real-world performance. We believe the way to do this is through validating the process. The research project for doing this would involve creating an STE for a current system using the same process as is used to generate envisioned world STEs. Researchers would then pretend that the current world is only envisioned, and create an STE for a situation in which they have access to real-world performance. They would examine whether performance in the STE matches that in the real world. If so, then confidence in the results of envisioned world STEs produced with this process would be increased. If not, the discrepancies would have to be examined with care. In this situation it would be all too easy to "cook" the STE, unintentionally or not, so that it was merely a tautological restatement of the existing real-world performance rather than a hypothesis about an unknown. The most straightforward way this would happen would be if existing SMEs were interviewed and then the information was compiled into the STE so that it was merely a description of what they do. Running those same SMEs through the STE would, by definition, result in complete correlation with real-world performance, and would fail completely in helping us gain confidence in our envisioned world STEs. To guard against this we must proceed with the utmost rigor and openness, ensuring that we internally review our execution of the process and secure extensive impartial external review as well. Under these conditions of scrutiny we can be confident of a valid and reliable result.

In this paper we have presented a process for designing, developing, and using STEs that provides a level of detail not previously published. However, there is still further refinement to be made to this process. Every activity in the process can benefit from further experience reflected back to the research and practitioner communities. One area we are focusing on is the use of the element matrices in creating the abstract design. Further experience will allow us to say more about how different kinds of data are analyzed into elements and how those elements are captured in the matrices. In future efforts we will evaluate the matrices to determine if they are capturing the necessary and sufficient information needed to create quality lightweight STEs. In our experience designing STEs we have noted a similarity to object-oriented design of software systems and simulations. Reviewing the literature on object-oriented design methods, for example Wirfs-Brock, Wilkerson, and Wiener (1990), may provide additional insight into techniques for STE design.

Another area for investigation centers on the ability of practitioners to create reliable and valid STE experiments. One of our goals in this line of research is to create a robust lightweight STE design and experimentation method that those with minimal formal training in experimentation and experimental design could use to conduct useful investigations. We believe that lightweight STEs address a pressing need in the practitioner community. Additional work needs to be done to understand the parameters within which practitioners can create and use STEs. At the current level of development, we could expect that practitioners could create derivative STEs from an exemplar created by an experienced STE designer and experimenter. Further work is needed to create guidelines for how that might be done in a safe manner, and to make the STE development process more robust so there are fewer judgments based in large measure on tacit expertise in experimentation.

8. Conclusion

Developing STEs is much like developing any other sort of product. We can use much of what is known about product development to make the creation of STEs more efficient and effective. However, STEs are different from other sorts of products. We therefore extend and modify existing practices to support the unique demands of using STEs in force transformation experimentation.

We have defined lightweight STEs for applied research in force transformation as minimalist experimental settings that retain cognitive authenticity. We have set forward a process for creating STEs, and a framework for documenting their design. Further experience is needed to learn how to optimize the creation of STEs so that they can keep pace with force transformation efforts in the US Armed Forces. We believe that the process and framework described in this paper will create quality experimental settings. As scientists, we believe in “trust, but verify” as a core principle. To that end, we have proposed a plan for verifying that STEs developed in accordance with the guidance in this paper produce useful, valid, and reliable results.

Force transformation is a high stakes activity. As such, the activities guiding it should be called to a higher standard than those informing the creation of the next great e-commerce site. As identified in the future development section, the creation and use of lightweight STEs requires further study and development by the community of researchers working on force transformation.

In this paper we have outlined a process and tools for creating lightweight STEs. We have begun the development of an STE for exploring UAV coordination between the UE and UA in Future Force. This is a part of exploring Army Airspace Command and Control (A2C2) in Future Force. We have begun two campaigns. The first is to create a community of practice for the development and use of STEs as a lightweight adjunct to large-scale exercises and simulations. The second campaign is to bring the cognitive perspective to the conversation about A2C2 force transformation. We look forward to the next steps in both these campaigns.

9. Acknowledgements

This work was performed through participation in the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009.

10. References

Alberts, D. S., & Hayes, R. E. (2002). *Code of best practices for experimentation*, from http://ww.dodccrp.org/publications/pdf/Alberts_Experimentations.pdf

Brehmer, B. (2004). Some protection on microworlds. In S. G. Schiflett, L. R. Elliott, E. Salas & M. D. Coovert (Eds.), *Scaled worlds: Development, validation, and applications* (pp. 22-36). Burlington, VT: Ashgate.

Cooke, N. J., & Shope, S. M. (2004). Designing a synthetic task environment. In S. G. Schiflett, L. R. Elliott, E. Salas & M. D. Coovert (Eds.), *Scaled world: Development, validation, and applications* (pp. 261-278). Burlington, VT: Ashgate.

Flach, J. M., Jacques, P., Patrick, D., Amelink, M., van Paassen, R., & Mulder, M. (2003). A search for meaning: A case study of the approach-to-landing. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 171-191). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly, 2*, 205-227.

Gugerty, L. (2004). Using cognitive task analysis to design multiple synthetic tasks for uninhabited aerial vehicle operation. In S. D. Schiflett, L. R. Elliott, E. Salas & M. D. Coovert (Eds.), *Scaled worlds: Development, validation, and applications* (pp. 240-262). Burlington, VT: Ashgate.

Gugerty, L., DeBoom, D., Walker, R., & Burns, J. (1999). *Developing a simulated uninhabited aerial vehicle (UAV) task based on cognitive task analysis: Task analysis results and preliminary simulator performance data.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Chicago, IL.

Hardin, G. (1968). The tragedy of the commons. *American Association for the Advancement of Science, 162*(3859), 1243-1248.

Hess, S. M., MacMillan, J., Elliott, L. R., & Schiflett, S. (1999). Team-in-the-loop, synthetic simulation: Bridging the gap between laboratory and field research. *Proceedings of the Human Factors and Ergonomics Society, 43rd Annual Meeting, 1*, 308-312.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems, 18*(3), 81-85.

Martin, E., Lyon, D. R., & Schreiber, B. T. (1998). Designing synthetic tasks for human factors research: An application to uninhabited air vehicles. *Proceedings of the Human Factors Society, 42nd Annual Meeting, 1*, 123-127.

Maulsby, D., Greenberg, S., & Mander, R. (1993). *Prototyping an intelligent agent through wizard of oz.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems.

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155-169.

Roesler, A., Feil, M., Woods, D. D., Puskeiler, A., & Tinapple, D. (2001). *Animock: Design tells (shares) stories about the future*, from http://csel.eng.ohio-state.edu/woods/practice-centered/CSELanimock_web.pdf

Ross, K. G., & Pierce, L. G. (2000). Cognitive engineering of training for adaptive battlefield thinking. In *IEA14th Triennial Congress and HFES 44th Annual Meeting* (Vol. 2, pp. 410-413). Santa Monica, CA: Human Factors.

Wirfs-Brock, R., Wilkerson, B., & Wiener, L. (1990). *Designing object-oriented software*: Prentice Hall.