DRAFT
EXPERIMENTATION TYPOLOGY

"A fool ... is a man who has never tried an experiment in his life." Erasmus Darwin

Background

The term "experiment" has become a major "buzz word" in the Department of Defense over the past two years. Activities that would have been termed assessments, evaluations, proofs of concept, tests, demonstrations, and even exercises in past years are suddenly categorized as experiments. This development has been encouraged by reform oriented members of the national security community, including influential members of Congress, who are concerned that the United States military may be lulled into a false sense of security by the absence of peer force threats. This concern is particularly high during an era when weapons of mass destruction and new technologies (particularly information technologies) create revolutionary opportunities for transforming warfare and the key arenas command and control, from sensing and fusion through decision making and battle management.

Experimentation has become seen as good within DoD because it is associated with science and new technologies, but also because, in the words of senior officers, experimentation allows, even benefits from, failure. Classic military exercises cannot be allowed to fail because they have an important role in training and because their failure reflects badly on the participating commanders and staffs, with potential career implications. Tests, assessments, and evaluations; particularly those focused on new systems and equipment, cannot be allowed to fail because they represent the culmination of lengthy and expensive research and development programs, and because their failure also has implications for the responsible people and organizations. Demonstrations are efforts to showcase new technologies and systems, so their failure defeats their purpose. Moreover, technology demonstrations have increasingly become a route around complex and cumbersome formal test and evaluation programs that are expected to leave behind systems improve that military capability in the field. By contrast, experimentation is seen as a legitimate "voyage of discovery" and a relatively systematic way to explore new approaches and the potential of new systems. Hence, experiments are seen as relatively risk free for both organizations and individuals and therefore attractive both to innovators and the military organizations asked to accept the risks associated with innovation.

Types of Experiments

Relatively few of those embracing the concept of experimentation within the Department of Defense have paid serious attention to the underlying concept. In fact, the term arises form the Latin, *experiri*, to try. Experimental knowledge differs from other knowledge in that it is always founded upon experience and observation. In other words, experiments are always empirical. A formal definition of experiment is, "A test made to demonstrate a known truth, to examine the validity of an hypothesis, or to determine the efficacy of

something previously untried." Indeed, all three of these meanings are relevant to DoD experimental activities in the recent past and planned for the future. Moreover, these three groups correctly distinguish the three major roles that DoD organizations have assigned to experimentation:

- **Hypothesis generation** experiments involve providing new systems and technologies in a setting where their use can be observed and catalogued. The idea is to simultaneously find out if the innovation is useful (enhances military capability) and how it can be employed. This application is similar to the old process in which new military hardware (aircraft, tanks, etc.) were developed against a set of technical specifications (fly faster, fly higher, turn faster, etc.), then given to a technical user community (the Army's boards, Air Force test organizations) where the tactics, techniques and procedures for effective employment could be worked out. In these applications, the goal is to identify apparent military benefits and develop systematic theories about the best way the new technology or system can be employed, which includes specifying the conditions under which it can be used (and their limits) as well as the results that can be expected. The results of these efforts were "theories" in that they were not considered validated until the weapons systems had been turned over to end users (fighting forces) and employed under field conditions. Similarly, hypothesis generation experiments usually occur early in the development cycle and will not normally provide enough information (or evidence) to conclude that the observed relationship is valid or will occur reliably. Hence, they will normally be followed by other experimentation and related activities designed to refine the knowledge gained and provide added reliability and validity.

- **Hypothesis testing** experiments are analogous to the classic efforts of scholars to advance knowledge by seeking to falsify specific hypotheses (if...then statements), whole theories (systems of related hypotheses that "explain" some area of inquiry or domain of knowledge), or observable hypotheses deduced from such a theory. These empirical experiments are efforts to build knowledge. That is, the experimenter(s) create a situation in which one or more factor(s) of interest (at the data level, dependent variables) can be observed systematically (measured) under conditions that vary the values of factors thought to cause change (independent variables) in the factor(s) of interest, while other potentially relevant factors (control variables) are held constant, either empirically or through statistical manipulation. Hence, experimental results in science are always caveated with *ceteris paribus,* or "all other things being equal." Since the numbers of causal factors and dependent variables of interest in the military arena are both very large, a great deal of hypothesis testing experimentation is implied when military innovation is attempted. Considerable thought and effort will be required to plan sound experimental programs, both to ensure that individual experiments generate useful results and to ensure that large programs are designed to accumulate knowledge systematically and effectively across multiple experiments.

- **Demonstration of known truth** is analogous to experiments conducted in high school and college laboratories, where the students follow instructions that allow

them to demonstrate, for themselves, that the laws of chemistry and physics operate as the underlying theories predict. These are similar to the technology demonstrations that have become significant within DoD in the past several years. However, the key difference is that demonstrations conducted as part of an experimental program will require systematic data collection in order to document the impact of new systems and technologies. Hence, they quantify the results that demonstrate the expected impacts while, at the same time, both validating the earlier research and experimentation and helping to establish a baseline against which the impact of future innovations can be measured.

All three types of experimentation have already taken place within DoD and are part of the overall planning for future experimentation at the Service and Joint levels. Hence, they properly belong to different parts of the research, development, and innovation process. Hypothesis generation experiments should ideally be conducted when the uses and limits of innovations intended to create new military capabilities are being explored. They should both provide indications of the potential utility of the innovation and also help to identify the best way(s) to employ them and the non-technological changes (doctrine, staffing, training, etc.) needed to permit full benefit from the innovation. Hypothesis testing experiments explore the dynamics of the innovation and the changes it enables or forces in the dynamics of the C2 process. They are about cause and effect and establishing valid and reliable knowledge about the uses, limits, unintended consequences, and benefits available from the innovation. The bulk of experimental effort should, in an ideal DoD program, be spent on these hypothesis testing efforts. Demonstration experiments should occur only when the dynamics and benefits of a particular innovation (or set of related innovations) have been established. Their primary purpose is to demonstrate the efficacy of the innovation to the user or operational community. They will differ from technical demonstrations only in the rigor with which they are observed and the benefits from the innovation are measured. This rigor, however, is important in that it provides the kind of evidence that can guide budgetary decisions and establishes performance baselines against which future innovation can be measures. For some parts of the user community, but by no means all, this empirical evidence will help to establish the credibility of the innovation.

Given that these three types of experimentation have somewhat different objective and roles within the DoD research, development, and acquisition processes, they should be managed somewhat differently. For example, the product of hypothesis generation experimentation should be a set of hypotheses which are considered important to explore further, but do not stand as established knowledge. Hence, they will often be conducted with less rigor (and therefore lower costs) and across a broader range of contexts than the other two types. Hypothesis testing experiments, by contrast, require somewhat greater control and should yield structured data, information, and knowledge. Moreover, the complexity of establishing cause and effect relationships with enough rigor to support investment decisions will require both designs that create layers of experimentation (mini-experiments inside experiments, experiments inside mega-experiments, linkages over time across sets of experiments) and also provide for sampling of the experimental

space in ways that allow credible inference across the range of important military threats, missions, and operating environments.

However, innovation within the Department of Defense is not a scientific endeavor with the luxury of infinite time to develop new knowledge. Information technologies are changing at an incredible rate. Adversaries have the potential to leapfrog generations of systems and technologies and can be anticipated to adapt commercial innovations to military applications. Moreover, the number and variety of opportunities to conduct experiments, particularly outside the training exercises which must use existing doctrine, organizations, and personnel, will continue to be modest. Hence, DoD's experimentation program will sometimes mix the three types of experimentation. In its simplest form, this may involve efforts to collapse the process by merging hypothesis generation and hypothesis testing experiments. A more complex variation may involve using hypothesis testing experiments on one system or technology to provide the context for hypothesis generation experiments on different innovations. While such combinations are certainly possible, they must be conducted carefully to ensure that each effort generates valid and reliable results to the research issues. Failure to keep the different goals and products well in hand will almost certainly lead to ambiguous results and the need for costly repetitions.