CHAPTER SIXTEEN

Synthetic Task Environments: Measuring Macrocognition

John M. Flach, Daniel Schwartz, April Bennett, Kyle Behymer, and Wayne Shebilske

Introduction

*I claim that many patterns of Nature are so irregular and fragmented, that,

compared with Euclid – a term used in this work to denote all of standard
graphics – Nature exhibits not simply a higher degree but an altogether different
level of complexity....

The existence of these patterns challenges us to study those forms that Euclid
leaves aside as being “formless,” to investigate the morphology of the
“amorphous.” Mathematicians have disdained this challenge, however, and
have increasingly chosen to flee from nature by devising theories unrelated to
anything we can see or feel. (Mandelbrot, 1983, p. 1)

Consistent with Mandelbrot’s comments comparing the “cold” geometry of Euclid with
the patterns of nature, there seems to be a growing dissatisfaction with the ability of classical
experimental approaches to human information processing to capture the complexity of
cognition “in the wild” (e.g., Hutchins, 1995). While many cognitive scientists have “disdained
this challenge” and have fled from the apparently amorphous patterns of everyday work to
study sterile, logical puzzles, a few have been plagued by a nagging fear that this research may
not be representative of how people experience life. In some sense, the construct of
Macrocognition reflects the challenge to address the patterns of cognition as they appear in
natural work contexts. In addressing this challenge, questions are raised about the very nature
of cognition and thus about the appropriate ways to measure it. There is a growing consensus that the cold geometry of micro laboratory tasks (e.g., reaction time, tracking, logical puzzles) is not capturing the complexities of human performance in natural settings.

This chapter first considers how assumptions about cognition from a “macro” perspective may differ from assumptions that have been made from “micro” perspectives. In doing this, we raise important questions about measurement. The chapter concludes by considering the use of synthetic task environments as one means for wrestling with both the theoretical assumptions and the measurement challenges of a “macro” approach to cognition.

**What’s the system?**

One of the first decisions researchers must make is to identify the phenomenon or system of interest. That is, they must identify what dimensions are endogenous to the phenomena – essentially the “state dimensions.”

As the following quote from Marr (1982) suggests, from the start researchers interested in humans as information processors, recognized that the system must include both the agent and the task demands:

... *the critical point is that understanding computers is different from understanding computations. To understand a computer, one has to study the computer. To understand an information-processing task, one has to study that information-processing task. To understand fully a particular machine carrying out a particular information-processing task, one has to do both things. Neither alone will suffice.* (Marr, 1982, p.5)
This suggests that the system must be defined in a way that includes both properties of the cognitive agents (i.e., constraints on awareness) and properties of the task or problem space (i.e., constraints on situations). Although the pioneers in the field (e.g., Marr, 1982; Newell & Simon, 1972) understood this, one gets a distinct impression from the research literature on human cognition and its application in terms of human factors that the task or problem space is arbitrary. This literature creates a distinct impression that the system of interest is “in the human’s head.” And that the research goal is to characterize the internal “limitations” within isolated stages of information processing so that these limitations can be accommodated in the design of complex systems. The primary motivation for choosing one laboratory task or another is the ability to isolate specific stages of processing (e.g., encoding or decision making) or specific internal constraints (e.g., a bottleneck or resource limit). There is little discussion of how well such tasks represent the demands of natural situations.

Thus, an important difference between micro- and macro-approaches to cognition is that microcognition defines the system as a process internal to an agent, whereas macrocognition defines the system to include both the agent and the problem demands. In other words, macro approaches consider cognition to be a process that is “situated” or “distributed.” And it assumes that understanding the “situation” is a critical component to a computational theory of cognition. In fact, to a large extent the “computation” is an adaptation to the problem or situation constraints. The implication for measurement is that we need to consider how to measure both situations and awareness. And we must measure them in a way that we can index the fitness of one relative to the other. They are not two separate systems, but two facets of a single system. The computational system (i.e., the cognitive phenomenon) depends on the fit between awareness and situations.
Measuring Situations

When considering measuring situations, it is important to begin with fundamental lessons about the nature of information. The information value of an event (e.g., selecting the number 42 from a jar) cannot be determined unless the possibilities (e.g., the other numbers in the jar) are specified. In simple choice decision paradigms (e.g., Hicks, Hyman), the other possibilities are well-defined in terms of the number and probabilities of alternatives. However, how do you specify the possibilities when the task is controlling a nuclear power plant, flying a modern aircraft, or directing air operations during battle; much less when the question has to do with the life of meaning or the meaning of life (Adams, 1979)?

From the micro perspective on cognition, there is a natural tendency to extrapolate from the measures that provided well-defined performance functions in the laboratory. This has stimulated initiatives to reduce events in a nuclear power plant and other complex situations to probabilities and durations that can be integrated using linear techniques like THERP or Monte Carlo style simulations such as MicroSAINT.

From the macro perspective, there is great skepticism about whether the dynamics (or possibility space) of many natural situations can be captured using event probabilities or time based measures alone. The alternative is to describe the constraints that shape the space of possibilities in terms of goals/values, general physical laws, organizational constraints, and specific physical properties. This perspective is well illustrated by Rasmussen’s (1986) Abstraction Hierarchy. The Abstraction Hierarchy is significant as one of the first clear specifications of the different classes of constraints that limit possibilities in natural work environments. In relation to the focus of this chapter, it is important to recognize that
measurement is a form of abstraction. Thus, the Abstraction Hierarchy is a statement about measurement. In fact, we suggest that it could easily have been termed a Measurement Hierarchy, where each level suggests different ways to index constraints within the work domain.

In the context of specifying or measuring situations, the Abstraction Hierarchy provides a useful guide for thinking about the various levels or types of constraints that shape the field of possibilities within a work domain. And it stimulates discussion about relations within and across these levels. Thus, when considering how to measure situations, one must consider both the need to describe the constraints in ways that reflect significant relations both within and across the various levels. Vicente (1999) illustrates this very clearly with the DUal REServoir System (Duress) example of a process control micro-world simulation. One caution is that while Duress is a great illustration – it represents a task with fairly well defined goals and constraints -- in many domains the constraints will be far more amorphous and discovering the right metrics to characterize the significant relations among the constraints is a significant challenge. But it is a challenge that must be engaged if there is to be any hope of understanding the computations involved in cognitive work.

Let’s consider some of the levels associated with situation constraints and the issues associated with measuring them. First, consider goals and values. Even in micro laboratory tasks (e.g., reaction time or signal detection) it is evident that tradeoffs between goals (e.g., speed versus accuracy or hits versus false alarms) have a significant role in shaping performance. Natural work domains are typically characterized by multiple goals and success often depends on balancing the demands associated with these goals (e.g., setting priorities or precedence). An important question for measurement is how to index performance with respect

to multiple goals so that the data can be integrated across the goal dimensions in a way that will reflect whether performance is satisfactory with regard to the aspirations for the system. Measures should allow some classification (satisfactory vs. unsatisfactory) or ordering (better or worse) of performance with respect to these aspirations.

Brungess’ (1994) analysis of the Suppression of Enemy Air Defenses (SEAD) is an important example of someone who is explicitly wrestling with the problem of how to measure performance relative to goal constraints. For example, he writes:

*SEAD effectiveness in Vietnam was measured by counting destroyed SAM sites and radars. Applying that same criteria to SEAD technologies as used in Desert Storm yields a confused, possibly irrelevant picture. SEAD weapons and tactics evolution has outpaced the development of criteria to measure SEAD’s total contribution to combat* (p. 51 – 52).

At another level, it should be quite obvious how general physical laws (e.g., thermodynamics or laws of motion) provide valuable insight into the possibilities of natural processes (e.g., feedwater regulation or vehicle control). These laws suggest what variables are important for specifying the state of the system (e.g., mass, energy, position, velocity). These variables are critical both to the cognitive scientist interested in describing the computational processes and to the active control agents (whether human or automated) in terms of feedback (i.e., observability and controllability). Note that for the variables to be useful in terms of feedback, they must be indexed in relation to both goals and control actions. That is, it must be possible to compare the information fed back about the current (and possibly future) states with information about the goals, in a way that specifies the appropriate actions. Thus, questions about controllability and observability require indexes that relate goals, process states, and
controls (e.g., Flach, Smith, Stanard, & Dittman, 2004). This is a clear indication of the need to choose measures that reflect relations within and across levels of the Measurement Hierarchy.

In addition to considering relations across levels in the Situation Measurement Hierarchy, it is important not to lose sight of the fact that the measures should also help to reveal important relations to constraints on awareness. Remember the system of interest includes both the problem and the problem solver. Understanding general physical laws can be very important in this respect, because these laws suggest ways to organize information to allow humans with limited working memory to “chunk” multiple measures into a meaningful unit. This is a recurrent theme in the design of “ecological” displays – to use geometric relations to specify constraints (e.g., physical laws) that govern relations among state variables (e.g., Vicente, 1999; Amelink, Mulder, van Paassan, & Flach, 2005).

For the purposes of illustrating the general theme of measuring the situation using indexes that reveal relations both across levels in the hierarchy and in relation to constraints on awareness, we only describe this one level in the Situation Measurement Hierarchy. For more discussion of other levels in the abstraction/measurement hierarchy see Flach, Mulder, and van Paassan (2004).

Measuring Awareness

Micro approaches to cognition tend to focus on identifying awareness dimensions that are independent of (or invariant) across situations. Thus, they tend to address issues such as perceptual thresholds, memory capacities, bandwidths or bottlenecks, and resource limits – as attributes of an internal information processing mechanism. However, within this research...
literature, it is not difficult to find research that attests to the adaptive capacity of humans. For example, basic work on signal detection suggests that performance is relatively malleable as a function of the larger task context (e.g., expectancies and values). Even the sensitivity parameter (d’) is defined relative to signal and noise distributions. So, there is ample reason for skepticism about whether any attribute of human performance can be specified independently from the larger task context.

Whether or not it is possible to characterize constraints on human information processing that are independent of the task context, few can argue that humans are incredibly adaptive in their ability to meet the demands of natural situations. Macro-approaches tend to focus on this adaptive capacity and this raises the question of measuring awareness relative to a domain (i.e., skill or expertise). There is little evidence to support the common belief that expert performance reflects innate talent. Rather the evidence suggests that expert performance reflects skills acquired through extended, deliberate practice in a specific domain (Erikson & Charness, 1994). In fact, Erikson & Charness (1994) conclude that “acquired skill can allow experts to circumvent basic capacity limits of short-term memory and of the speed of basic reactions, making potential limits irrelevant” (p. 731).

Again, we feel that credit goes to Rasmussen (1986) as one of the first to explicitly recognize the flexibility of human information processing and to introduce a conceptual framework specifying important distinctions that must be addressed by any program to quantify human performance in natural contexts (i.e., the decision ladder and the SRK distinction between Skill-, Rule-, and Knowledge-based processing). The decision ladder explicitly represents the shortcuts that might allow experts to ‘circumvent’ basic capacity limitations. And the SRK distinction provides a semiotic basis for relating the properties of the
situation (e.g., consistent mapping) to the potential for utilizing the various shortcuts (Flach & Rasmussen, 2000).

Rasmussen (1986) illustrates how the decision ladder can be utilized to visualize qualitatively different strategies for fault diagnoses. This is clearly an important form of measurement that helps to index performance in relation to potential internal constraints on awareness and in relation to the demands of situations. Furthermore, it allows these strategies to be compared to normative models of diagnoses.

Consistent with the basic theory of information, it is important not to simply ask what experts ‘know’ and what strategy experts typically use; we must explore the possibilities about what experts ‘could know’ and about what strategies might be effective in a given situation. Note that the awareness of experts will be constrained by the types of representations that they have been exposed to. For example pilots and aeronautical engineers utilize very different forms of representations for thinking about flying. Thus, there can be striking contrasts for how these different experts think about flight performance. Exploring the differing forms of awareness can be important for differentiating more and less productive ways for thinking about a problem such as landing safely (Flach, Jacques, Patrick, Amelink, van Paassan, & Mulder, 2003). It is important to keep in mind that the best operators (e.g., pilots or athletes) often do not have the best explanations for how and why they do what they do.

Considering alternative representations across different experts can suggest possibilities for shaping ‘awareness’ through the design of interfaces. As Hutchins’ (1995) work clearly illustrates, the choice of a specific technique for projecting the world onto the surface of a map has important implications for the cognitive processes involved in navigation. Again, this is a key theme behind the construct of ecological interface design – to shape the nature of
awareness to facilitate information processing (in terms of allowing shortcuts and supporting multiple pathways to satisfactory outcomes) (Rasmussen & Vicente, 1989; Vicente & Rasmussen, 1990). The point is not to simply match existing mental models, but to design representations that help shape the mental models to enhance awareness and resilience.

**Measuring Performance**

At the end of the day, one of the most important measurement challenges is to be able to index the quality of performance. For example, to be able to index whether one interface, training protocol, incentive system, leadership style, organization plan, etc. leads to ‘better’ performance than another. Micro-approaches to cognition prefer to focus on one dominant measure (e.g., time to completion or percent correct) to index quality. Even when there is clear evidence of the potential for tradeoffs (e.g., speed versus accuracy), micro-style research tends to frame the task to clearly emphasize one dimension (e.g., “go as fast as possible with zero errors”). These approaches typically assume that performance functions are monotonic (e.g., “faster is better”). This is typically generalized to research in human factors, where two designs might be evaluated in terms of which design produces a statistically significant advantage in response time. But whether a statistically significant difference in response time leads to a practical gain in work performance is difficult to address with the Micro-approach. The General or Chief Executive Officer (CEO) who asks whether the ‘improved’ system will be worth the cost in terms of achieving the objectives that are important to him (e.g., greater safety or a competitive advantage) rarely gets a satisfactory answer.

In the everyday world, there is rarely a single index of satisfaction. As discussed in the section on measuring situations, there are typically multiple goals that must be balanced – a
good system should be fast, accurate, safe, and not too expensive. This requires either multiple performance measures or at least an explicit integration of indexes associated with the various goals to yield a single ‘score’ for ranking goodness or at least for distinguishing between satisfactory and unsatisfactory performance. And rarely are the quality indices monotonic. That is, success typically depends on responding at the right time (not too early or too late). At least for closed-loop systems there will always be a stability boundary that limits the speed (i.e., gain) of response to stimuli. Thus, response speed is rarely monotonic – a system that is too fast can become unstable (e.g., pilot induced oscillations).

It is impossible to address questions about the right information, the right place, the right person, or the right time without considering the specific problem that is being solved (i.e., the work domain or task). ‘Right’ is highly context dependent. It cannot be addressed by a micro-research program that is designed to be context independent. This is an important motivation for a macro-approach to cognition and work – to specify the specific criteria for satisfying the demands of specific work domains.

In order to know whether a difference in response time is practically significant, it can be useful to compare this against landmarks that reflect the optimal or best case situation. Here is where analytic control models (e.g., the optimal control model) or Monte Carlo simulations (e.g., Microsaint) can be very useful. Not as ‘models’ of human information processes, but as ways to explore the boundary conditions of performance. What is the best possible performance assuming certain types of processes? How do changes at one step in a process (e.g., to speed or accuracy) or in properties of a sensor (e.g., signal-to-noise ratio) impact system performance? Where are the stability limits? In this sense, the models are being used to explore boundaries (or limits) in the workspace. These boundaries may provide important

insights into what are realistic targets for improvement and into the practical value of specific improvements. In essence, these models can suggest ‘normative’ landmarks against which to assess actual performance.

An example of a case where this type of insight might be useful is a recent initiative on the part of the Air Force to reduce the response time for executing dynamic targets to single digit minutes. This is motivated by the threat of mobile missile systems (e.g., Scuds) that can fire a missile and then move to cover within about 10 minutes. Few have raised the question about whether this response time is realistic given the unavoidable lags associated with acquiring the necessary information and communicating with the weapons systems. We fear that the blind pursuit of this single digit minute goal may lead to instabilities and unsatisfactory solutions to the overall goals of the Air Force. Rather than “reacting” faster, the solution to the Scud missile problem may depend on improving the ability to predict or anticipate launches (e.g., Marzolf, 2004). Thus, the solution to Scud missiles may rest with the design of the air battle plan, to include dedicated aircraft to patrol areas where launchers are likely to be hidden. This solution does not require speeding the dynamic targeting process for dealing with events not anticipated in the air battle plan.

Another important consideration for measuring performance is the distinction between process and outcome. In complex environments, an optimal process can still result in a negative outcome due to chance factors that may be completely beyond control. For example, a coach can call the perfect play that results in a touchdown and have it nullified by a penalty flag, incorrectly thrown by a poor referee. Thus, it is important to include measures of process as well as measures of outcome. And it is important to have standards for measuring process as well as outcome. For example, most military organizations have doctrine that provides
important standards for how processes should be conducted (see Chapter 8 for an example on communicating commander’s intent to subordinates).

The astute reader should realize that as we talk about performance measurement, we are covering some of the same ground that was discussed in terms of measuring situations. We are talking about ends (goals and values) and means (processes). In classical Micro-approaches that define the system of interest as inside the cognitive agent, the situation is typically treated as an independent variable and performance measures are treated as dependent variables. This creates the impression that these are different kinds of things. However, in natural work ecologies, understanding the situation requires consideration of both means and ends. So, using the Abstraction Hierarchy to think about situations will go a long way toward addressing questions about performance measures. And it should help to frame these questions in terms that are meaningful to the problem owners (those who have a stake in success).

Thus, consistent with our discussion about situations, we believe that it is useful to think about a nested hierarchy of performance measures, where higher levels in the hierarchy reflect global criteria for success (e.g., how do you know whether you are winning or losing); and where lower levels address subgoals and process measures that reflect the means to higher level goals (e.g., showing patterns of communication or organization). It is important to keep in mind that the primary goal of measurement is to reveal the patterns of association between process and outcome. In other words, a key objective is to connect the micro-structure associated with the design and organization of work activities to qualitative changes associated with global indexes of quality!

**Synthetic Task Environments**

In contrasting micro- and macro-approaches to cognition, our intent is not to eliminate micro-level research, but rather to make the case that this is only one element of a comprehensive research program. We fear that a research program that is exclusively framed in terms of low dimensional tasks will not satisfy our goals to understand cognition in natural contexts or to inform the design of tools to support cognitive work. Thus, the goal of a macro-approach is to enrich the coupling between the laboratory and the natural world. It is in this context that we would like to suggest that research employing synthetic task environments can be an important means for bridging the gap between basic experimental research and natural cognition.

We use the term synthetic task environment to represent experimental situations where there is an explicit effort to represent the constraints of a natural work domain. This is in contrast to experimental paradigms designed around the parameters of a particular analytic model (e.g., choice reaction time or compensatory tracking) or designed to isolate a specific stage of an internal process (e.g., visual and memory search tasks). It is also in contrast to micro-world research that attempts to represent the complexity of natural domains, without representing the constraints of specific actual domains (e.g., space fortress or other research using computer games). In a synthetic task, the work domain has to be more than a cover story. The ‘task’ must be representative of some natural work – even though the implementation is synthetic, typically utilizing a simulation.

For example, research using a flight simulator may or may not satisfy our definition for synthetic task research. If the focus is on flight performance, perhaps in relation to a specific training protocol, to compare alternative interfaces, or to evaluate different procedures, then this is consistent with our definition of synthetic task research. However, if the focus is on
cognitive workload and the flight task is just one aspect of a multiple task battery, then we
would not consider this to be synthetic task research. Again, this does not mean that such
research is not valuable, but we simply want to emphasize that for synthetic task research, the
focus should be on the impact of specific natural constraints of the work on cognitive
processes. The key to synthetic task research is NOT the use of a simulator, but the framing of
research questions with respect to properties of the natural task or the natural work domain!

A second important facet of synthetic task environments is the ability to measure
performance at multiple levels, as discussed in previous sections of this chapter. The synthetic
task environment should allow performance to be scored relative to global objectives (e.g., was
a landing successful; was the mission completed successfully). And it also should allow direct
measures of the work processes (e.g., time history of control and communication activities and
of system state, such as the actual flight path).

One of the important goals for synthetic task research is to provide empirical data with
respect to the coupling of global metrics (goals and values) and micro metrics (work activities
and movement through the state space). The goal is to empirically relate variations at one level
of the measurement hierarchy to variations at the other levels. Thus, for example, this may
allow the question about whether a significant difference in reaction time is practically
significant with respect to the global intentions for the system to be addressed empirically. Do
quantitative differences in response time to a particular class of events lead to increased
probability of successfully completing the mission?

Another consideration with respect to synthetic task research is the question of fidelity.
How much is enough? This is a bit tricky, because this is one of the questions that we are
asking when we frame questions around situations. What are the important constraints and how
do they interact to shape performance? For this reason, the issue of fidelity can only be addressed iteratively. In general, it is best to start with as much fidelity as you can practically afford and to assume that it is not enough! The performance observed in synthetic tasks needs to be skeptically evaluated relative to generalizations to natural domains. In our view, to be effective, a program of synthetic task research should be tightly coupled to naturalistic field studies. The patterns observed in the laboratory need to be compared to patterns observed in naturalistic settings. In this way, it may be possible to titrate down to identify critical constraints and interactions. The synthetic task observations allow more rigorous control and more precise measurement. But there is always the possibility that the patterns observed in the synthetic task are a result of your simulation and that they are not representative of the natural domain of interest. Ideally, however, synthetic task environments can improve our ability to see and quantify patterns during more naturalistic observations.

It is also important to note that questions of fidelity should not be framed simply in terms of the simulation device. Consideration must be given to the participants of the research. Are they representative of the people who do this work in natural settings, in terms of knowledge, skill, motivation, etc.? Consideration also must be given to the task scenarios. Are the tasks representative of the work in the natural context in terms of probability of events, consequences, and organization? (For more on the design of complex task scenarios, see Chapters 13-15) But more importantly, are the experiences of the participants representative of experiences in the real work domain (e.g., in terms of stress)?

In order to bridge the gap between laboratory research and cognition in the wild, synthetic task research will be most effective when the questions are driven by field observations of natural environments and when the multiple nested measures are motivated by
1) the values of the problem owners, 2) by normative models of the work (e.g., information theory, control theory, queuing theory), and 3) by basic theories of cognition. Currently, each of these three motivations has its champions and there seems to be a debate over which of these motivations is optimal. In our view, all three motivations are critical and none of these motivations alone will meet our aspirations for a science of cognition. With respect to these three motivations, the synthetic task environment may provide a common ground to facilitate more productive coordination between the disparate constituencies across the basic and applied fields of cognitive science.

Finally, it is important to recognize the inherent limits on any controlled scientific observation. The results will depend in part on properties of the phenomenon of interest and in part on the choices we make in designing the synthetic task environment. It is important to resist the temptation to become infatuated with a particular experimental paradigm (whether micro-task or specific synthetic task environment). It is important to leave ultimate control to Nature!

**Conclusion**

_Nature does exist apart from Man, and anyone who gives too much weight to any specific [ruler] … lets the study of Nature be dominated by Man, either through his typical yardstick size or his highly variable technical reach. If coastlines are ever to become an object of scientific inquiry, the uncertainty concerning their lengths cannot be legislated away. In one manner or another, the concept of geographic length is not as inoffensive as it seems. It is not entirely “objective.”_ [Mandelbrot, 1983, p. 27]

The quote from Mandelbrot reflects the difficulty in measuring a natural coastline – as the size of the ruler get smaller, the “length” of the coastline can grow to infinity. If a simple attribute like “length of a coastline” creates this difficulty for measurement, how much more difficult is the problem when the Nature that we are trying to measure involves Humans themselves. In our view, perhaps, this might be the biggest differentiator between micro- and macro-approaches to cognition. The micro-approach clings to the classical idea that it is possible to stand outside of ourselves to “objectively” measure cognition, work, or situation awareness. The macro-approach believes that this is a myth.

The macro-approach understands that measurement is not neutral. There is no privileged measure or privileged level of description! Every measure, every level of description, every perspective offers an opportunity to see some facet of Nature, but hides other facets. Thus, understanding requires multiple measures, multiple levels of description, and/or multiple perspectives. In this respect, the Abstraction Hierarchy or Measurement Hierarchy is simply a way to be explicit about the need to measure at multiple levels and a framework to guide the search for patterns that are invariant over multiple perspectives. In other words, the only way to eliminate or unconfound the invariant of a specific measurement perspective from an invariant of Nature is to measure from multiple perspectives. One is more confident in attributing an invariant to Nature, when that invariant is preserved over many changes of observation point.

Note that this is not a special requirement for studying Humans or Cognition. This will be true for any complex phenomenon in Nature (e.g., weather systems or coastlines). By complex, we simply mean a phenomenon that involves many interacting dimensions or degrees of freedom.
It is humbling to realize that Nature/Cognition cannot be reduced to reaction time and percent correct; to realize that the convenient measures (in terms of experimental control or in terms of analytic models) will not yield a complete picture; to realize that measures that work within the constraints of the ideals of Euclidean geometry do not do justice to the curves of natural coastlines. We get a distinct impression that the field of cognitive science is searching for a mythical holy grail – that is, a single framework (neuronets, neuroscience, chaos, etc.) and a specific measure (42, MRI, 1/f scaling, etc.) that will provide the key to the puzzle. We are skeptical.

Complex systems are difficult. They require multiple levels of measurement. An attractive feature of synthetic task environments is that they allow many measures (from micro-measures specifying the complete time histories of activity and state change, to macro-measures specifying achievement relative to the intentions of operators and system designers). The problem is making sense of all this data, weeding through the data to discover the patterns that allow insight, prediction, deeper understanding, and generalization. Success in this search requires the intuitions available from normative systems theory (e.g., information, signal-detection, and control theory, computational and normative logic, nonlinear systems and complexity theory), from controlled laboratory research, and from naturalistic field observations. Again, none of these perspectives on research is privileged. We expect that if there are answers to be discovered, they will be found at the intersection of these multiple perspectives. Thus, the value of synthetic task environments is to create common ground at the intersection of these various perspectives where we can constructively debate and test alternative hypotheses about the nature of cognitive systems.

Acknowledgments

We thank the organizers of the Measurement Workshop, Emily Patterson and Janet Miller, for giving us a forum to express our rather eccentric views about cognition, measurement, and the implications for system design. These views have evolved over many years of research. This research has benefited from several sponsors, most significantly the Air Force Office of Scientific Research (AFOSR) and the Japan Atomic Energy Research Institute (JAERI).

References


