CHAPTER 2

Current Objective: More Complete Performance

There are a wide range of behaviors that have yet to be incorporated into existing models. Included in this list are numerous additional relevant regularities about human behavior (see Boff & Lincoln, 1988, for a subset). The question that must be addressed is: which behaviors are the most important and most accessible to incorporate? We note here several of the most promising or necessary behaviors to be included next in models of human performance, based on our experiences and previous work.

The suggestions we make later tend to be based on modeling the individual. Much of the behavior being modeled currently in synthetic environments is different because it needs to include small and large groups and is aggregated across time or situations. As smaller time scales and more intricate and fine-grained simulations are developed and used, such as for modeling urban terrorism, the behavioral issues noted here will become more important.

We start with learning. While Pew and Mavor include learning as a useful aspect of performance, we believe learning to be essential. We also expand the case for including models of working memory, perception, emotions and behavioral moderators, and erroneous behavior. We then can examine higher-level aspects of behavior to be considered, starting with integration of models and ending with information overload.

2.1 Learning

Learning is mentioned as important in several ways by Pew and Mavor (1998). Learning (i.e., training) is the largest role of the military in peace time (i.e., rehearsal, p. 30), essential for multi-tasking behavior (pp. 114-115), an important aspect of human behavior (chap. 5), and important within groups (chap. 10). We cover learning again here.

Pew and Mavor mention several of the advantages of learning. There are several additional advantages that we can emphasize. Tactics are influenced by learning. In addition, there is a home-field advantage: working within your own territory, because you know it, makes additional tactics feasible and provides generally improved performance. (Working within your own territory would also provide some additional motivation.)

Including learning in models would provide a mechanism for producing different levels of behavior. Experienced troops, for example, would be different not in some numeric way in that they react faster (although this is probably true), but in a more qualitative way in that they know more and use different strategies. Learning modifies, constrains, and supports the use of computer interfaces (Rieman, Young, & Howes, 1996); similar effects may be found in exploring terrain and implementing tactics in new geographic spaces.

Programming—that is, creating the model directly—may be too difficult. It may be easier for models to learn behaviors than for these behaviors to be programmed directly. This argument has been put forward by connectionist researchers for some time.

Theoretical work in this area of learning has direct implications for training within the military and within schools. Models that learn can be used to understand and optimize learning (Ohlsson, 1992). If we can program models to learn, the behavior and knowledge that result may be different from the initial knowledge that the system started with or from the expert performance that we currently teach. This final knowledge may be useful for teaching. In the case of photocopying (Agre & Shrager, 1990), for example, better strategies arise through practice but are not valuable enough to teach. In military domains, it may be useful to find and then to teach the improved strategies that may arise from grossly extended practice, that is, tactics that are better but that no person has had enough practice to learn before. At that point, explanation of behavior will also become important to understand why the new behavior is useful so that it is trusted.

2.2 Expertise

Expert behavior has an important role to play in models of human performance (Shadbolt & O'Hara, 1997). One of the Western powers' greatest strengths is training in depth *and* breadth. Practice influences speed of processing and error rates, particularly under stress. If synthetic forces are to be used to test doctrine, the effect of training on expertise must be included.

Expert behavior has been studied extensively in recent years and a great deal is known about it (Chipman & Meyrowitz, 1993; Ericsson & Kintsch, 1995; Gobet, 1998; Gobet & Simon, 2000; Hoffman, Crandall, & Shadbolt, 1998). Some essential characteristics of expertise are highly developed perception for the domain material, selective search for solutions in that domain, and a good memory for domain-related material. In most domains, problem-solving behavior (search) differs as well: novices tend to search backward from the situation to find solutions and experts tend to search forward from the situation to find solutions (Larkin, McDermott, Simon, & Simon, 1980). Finally, transfer of expertise to other domains is limited.

Klein and his colleagues (e.g., Klein, 1997) have studied real-time performance in real settings (as opposed to laboratory settings) in detail, and have essentially found that the characteristics mentioned above are also critical in these situations. A number of rather extensive reviews have been undertaken of Klein's approach, which is often referred to as Naturalistic Decision Making (NDM) (e.g., Hoffman & Shadbolt, 1995). A method to elicit this type of knowledge has been developed by Klein and his associates. It is known as the Critical Decision Method and is described in Hoffman et al. (1998). The specifically real-time challenges of acquiring knowledge relating to perceptually cue-rich decision making are discussed in a second Defence Evaluation and Research Agency (DERA), United Kingdom, report by Hoffman and Shadbolt (1996).

Given the fact that it takes a long time to become an expert—the rule of 10 years or 10,000 hours of practice and study is often mentioned (e.g., Simon & Chase, 1973)—the size of the dataset has made it difficult indeed to study real-time learning on the road to expertise. However, real-time learning in simpler problem-solving tasks has been studied and modeling accounts have been provided (Anzai & Simon, 1979; John & Kieras, 1996; Nielsen & Kirsner, 1994; Ritter & Bibby, 2001). Some of these results may apply to expert learning in more complex tasks as well.

While experts vastly outperform non-experts in most domains, exceptions to this rule have been found in domains such as clinical diagnosis, clinical prediction, personnel selection, and actuarial predictions (Dawes, 1988). In these domains, experts perform only slightly better than non-experts, and typically perform worse than simple statistical methods, such as regression analysis. One other aspect of behavior that distinguishes experts from novices is the ability to recover from errors. An important question is to which category military diagnosing and prediction belong because of the uncertainties involved? And, based on this answer, what can be done (either by providing formal tools or by improving training) to remedy this situation and assist error recovery?

The effect of learning local environments and strategies (own and opponent's) must also be included. Having learned the local terrain probably explains much of the home-field advantage. How does this learning occur?

Within the sub-field of knowledge-engineering there have been considerable efforts to produce methodologies for the acquisition, modeling and implementation of knowledge-intensive tasks. It is a moot point whether the resulting decision-support systems are cognitively plausible. Nevertheless, these methodologies now provide powerful ways of constructing complex systems that exhibit task-oriented behavior. To this end, anyone engaged in engineering large-scale synthetic environments should look at the principles laid down in the most recent of this work. The most accessible source is probably Schreiber et al. (2000).

2.3 Working Memory

Central to all questions about human cognition and performance is the role of working memory. Working memory is implicated in almost all aspects of cognitive performance (Boff & Lincoln, 1986, Sec. 7; Just & Carpenter, 1992; Newell & Simon, 1972; Wickens, 1992). It is widely agreed that limitations of working memory are a major determinant of limitations of cognitive performance. Definitions of working memory are varied but for present purposes we can take it to refer to the mechanisms that maintain and provide access to information created or retrieved during the performance of a task.

Modern approaches to the psychological study of human working memory often take as their starting point the famous paper by Miller (1956) and argue that people can retain only around "7 +/- 2" items in short-term memory. Later work has tended to revise that estimate downwards, towards three to four items of unrelated information (Crowder, 1976; Simon, 1974).

A more recent and influential line of work by Baddeley (1986, 1997) presents working memory as a dual system for the rehearsal of information, consisting of (1) a phonological loop, that contains approximately 2 seconds of verbalizations, for the rehearsal of phonological, acoustic, or articulatory information (e.g., useful for repeating a phone number until you dial it); and (2) a visual-spatial scratchpad, with a smaller and less-determined capacity (e.g., useful when searching for an object that you have just seen), to play an analogous role for the maintenance of pictorial and spatial information.

Other approaches within experimental psychology place more emphasis on the role of working memory in both storing and manipulating temporary information (Daneman & Carpenter, 1980; Just & Carpenter, 1992). An important recent extension to the notion of working memory comes from the study of expertise, where Ericsson and Kintsch (1995) argue that after extensive practice in a particular domain people can, through specialized *retrieval structures,* use long-term memory for the rapid storage of temporary information (i.e., long-term working memory).

A recent book (Miyake & Shah, 1999) reviews a range of current approaches to the modeling of working memory, although many of the models do not have the explicitness and generality needed to support the simulation of human performance in complex tasks. Of those that do, their view of working memory varies widely. Some, such as ACT-R (Anderson & Lebiere, 1998) and CAPS (Just & Carpenter, 1992), consider working memory not as a separate structural entity but rather as an activated region of a larger, more general memory system, in which the limitations of working memory derive from a limited total quantity of activation. Just and Carpenter (1992), and more recently ACT-R models, have extended that view to the modeling of individual differences in working memory where different people are assumed to have different maximum quantities of available activation (Daily, Lovett, & Reder, 2001; Lovett, Daily, & Reder, 2000). A number of these ideas are put together by Byrne and Bovair (1997) who modeled (in CAPS) the way that a class of performance errors, in which people forget to complete subsidiary aspects of a task (such as removing the original from a photocopier), is affected by working memory load.

In contrast to these resource-limited models, Soar (Laird, Newell, & Rosenbloom, 1987; Newell, 1990) imposes no structural limitation on working memory. Using Soar, Young and Lewis (1999) explore the possibilities of working memory being constrained not by physical resources but by functional limitations and by specific kinds of similarity-based interference.

In summary, the current position is that human performance is known to be highly dependent on working memory and working memory load, and to be susceptible to factors such as individual differences (Just & Carpenter, 1992), distractions (Byrne & Bovair, 1997), emotion and stress (Boff & Lincoln, 1988), and expertise (Ericsson & Kintsch, 1995). Many existing models of human performance (e.g., as reviewed in Pew & Mavor, 1998) do not directly model the role of working memory. Models exist (Miyake & Shah, 1999), and some approaches to cognitive modeling (ACT-R, CAPS, Soar) have potential for improving predictions of human performance in realistic task situations by including more accurate theories of memory. There remains a need for the investigation and development of more explicit and complete models, with broader scope, of the role of working memory in human performance.

2.4 Emotions and Behavioral Moderators

Emotion, affect, motivation, and other behavioral moderators are increasingly being seen as factors that can and often do influence cognition. This view has received attention among a range of computer scientists and psychologists. Pew and Mavor (1998, chap. 9) lay out an initial case for including emotion as an internal moderator of behavior. The British HCI Group sponsored a one-day meeting on “Affective Computing: The Role of Emotion in

Human Computer Interaction" that attracted 70 people to University College, London (Monk, Sasse, & Crerar, 1999). Picard's (1997) recent book provides a useful review of emotions and computation in general. Sloman's (1999) review of the book and Picard's (1999) response are useful summaries. A further case is also made in the section on the Sim_Agent Toolkit. We present here an additional argument for including a model of emotions and behavioral moderators in models of synthetic forces, note two potential problems with existing models, and sketch an initial theory.

2.4.1 Further Uses of Emotions and Behavioral Moderators

Models of emotions and behavioral moderators may be necessary for modeling non-doctrinal performance such as insubordination, fatigue, errors, and mistakes. Many authors have also noted the role of emotion in fast, reactive systems (Picard, 1997, provides a useful overview). Individual differences in emotions may be related to personality and differences in problem solving. That is, the range of emotions may be best explained as an interaction that arises between task performance and situation assessment and an agent's likes, desires, and personal cognitive style. An argument is starting to be put forward that changes in motivation based on temporally local measures of success and failure may help problem solving (Belavkin, 2001; Belavkin & Ritter, 2000; Belavkin, Ritter, & Elliman, 1999).

2.4.2 Working Within a Cognitive Architecture

Emotions arise from structures related to cognition and should be closely related to and based on cognitive structures. All of the arguments for creating a unified theory of cognition (Anderson, Matessa, & Lebiere, 1998; Newell, 1990) also apply to creating a unified theory of emotion as well. The effects of emotions and other behavioral moderators on cognition are presumably not task-specific, so their implementation belongs in the architecture, not in the task knowledge.

Theories of emotions should thus be implemented within a cognitive architecture. This will allow them to realize all the advantages of being within a cognitive architecture, including being reusable and being compared to and incorporated within other models. Some models of emotions have been built within a cognitive architecture (Bartl & Dörner, 1998; Belavkin, Ritter, & Elliman, 1999; Franceschini, McBride, & Sheldon, 2001; Gratch & Marsella, 2001; R. Jones, 1998; Rosenbloom, 1998). Being created within an information-processing model has required them to be more specified than previous theories. Being part of a model that performs the task has also allowed them to make more predictions.

2.4.3 A Sketch of a Computational Theory of Emotions

An important aspect of cognition is to process sensory information, assign meaning to it, and then decide upon a plan of action in response. This is a real-time process in which new sensory information arrives continuously. This view is similar to the view put forward by Agre and Chapman (1987) about representationless thinking. The plan must therefore be dynamically reconfigurable and will often be abandoned in favor of a better plan midway through its execution. Elliman has a speculative view of the role of emotions in cognition, similar to Rasmussen's (1998) stepladder framework of behavior, which makes the following assumptions:

- The amount of sensory data available at any moment is too large for attention to be given to more than a small fraction of the data.
- The conscious consideration of the results of perception is an expensive process in terms of the load on neural hardware and also time-consuming.
- Most sensory processing is unconscious in its early stages in order that expensive conscious processes need consider only the *results* of perception. These results might include labeled objects with a position in space, for example "a tank moving its turret in that clump of trees." Conscious processes might well add further detail such as the type of tank and the range of its gun.
- Attentional mechanisms are needed to direct the limited high-level processing to the most *interesting* objects. These may be novel, brightly colored, fast-moving, or potentially threatening.
- Planning is an especially heavy computational process for the human mind and one that is difficult to carry out effectively under combat conditions. (Perhaps the best way to explain why military doctrine is useful is that it distills the best generic practice and trains the soldier to behave in a way that might well have been a chosen and planned behavior if the individual had the time and skill to formulate the action himself. The danger is that no doctrine can envisage all scenarios in advance and, on occasion, the use of doctrine in a rigid manner may be harmful.)
- From an evolutionary perspective this system of unconscious processing of sensory input, attentional mechanisms, and cognitive planning (together with speech-based communication) is a masterstroke of competence for survival. However, it has one crippling disadvantage—it is too slow to react to immediate and sudden attack.

Rapid reaction to possible threat without the time for much cognitive processing is clearly of huge value. In this framework emotion can be seen as kind of labeling process for sensory input. Fear particularly fits this pattern and is a label that causes selected sensory input to literally *scream* for attention. For this process to work rapidly it needs to be hardwired differently than higher-level cognitive processes. There is strong evidence that the amygdala is intimately involved in the perception of threat and able to trigger the familiar sensation of fear (e.g., Whalen, 1999). If this organ of the brain is damaged, individuals may find everyday events terrifying while not perceiving any need for alarm in life-threatening situations.

This rapid, emotive response to sensory data is relatively crude and prone to false alarms. Reactive behavior is triggered that may be involuntary, for example, the startle reaction and physiological changes due to the release of noradrenalin. After the reaction response, it takes time for cognitive processes to catch up and make a more informed assessment of the situation and actual threat. If this emotive, reactive stimulation is excited in a chronic manner then susceptible individuals may become less effective, with impaired ability to think and plan clearly. Any kind of anxiety is a form of stress. Because individuals have a finite capacity for absorbing it, excessive stress results in fatigue.

2.5 Errors

Ideally, military behavior is normative, that is, what *is* done is what *should* have been done. Human behavior does not always match the normative ideal of military behaviors. One of the most important aspects of human performance, which has often been overlooked in models of behavior and problem solving, is errors (although see, for example, Cacciabue, Decortis, Drozdowicz, Masson, & Nordvik, 1992; Freed & Remington, 2000; Freed, Shafto, & Remington, 1998). There is a consensus building about the definition of errors—for most people an error is something done that was not intended by the actor, that was not desired, and that placed the task/system beyond acceptable limits (e.g., Senders & Moray, 1991).

Part of the reason for omitting errors from models of behavior is the fallacy that they are produced by some special error-generating mechanism that can be bolted on to models once they are producing correct behavior on the task at hand. Often, however, the actions that precede errors would have been judged to be correct if the circumstances had been slightly different. In other words, as Mach (1905/1976) observed, knowledge and error both stem from the same source.

Evidence shows that novices and experienced personnel will often make the same errors when exposed to the same circumstances. The difference lies in the ability to notice and recover from these errors. Experienced personnel are more successful at mitigating errors before the full consequences arise. In other words, it is the management of errors that is important and needs to be trained (Frese & Altmann, 1989), rather than vainly trying to teach people how to prevent the inevitable.

2.5.1 Training About Errors

In any complex, dynamic environment, such as a military battlefield, the consequences of uncorrected errors are potentially disastrous. While normally a string of mistakes is required to create a disaster, the rapid pace of the battlefield and adversaries allows single mistakes to become more catastrophic.

There is, therefore, a real need to learn how to manage errors in an environment in which the consequences are less severe. An advantage of using synthetic environments is that comparative novices can experiment in unfamiliar situations, with restrictions approximating the real environment in time, space, enemy capabilities, and so on, but with the knowledge that the consequences of any errors can be recovered. In addition, multiple scenarios can be played out over a compressed time period, thereby providing the novice with a variety of experiences that would take many years to accumulate through exposure to situations in the real world. This can be a great training aid, literally giving years of experience in far less time. When novices were trained in aircraft electrical-system troubleshooting using a simulated system, they were able to acquire years of experience in months because the tutor let them practice just their diagnostic skills without practicing their disassembly skills (Lesgold, Lajoie, Bunzon, & Eggan, 1992).

2.5.2 Models That Make Errors

There are several process models complete enough to make errors, depending to some degree on the definition of error. Models that include errorful behavior exist in EPAM (Feigenbaum & Simon, 1984; Gobet & Simon, 2000), ACT-R (Anderson, Farrell, & Sauers, 1984; Anderson & Lebiere, 1998; Lebiere, Anderson, & Reder, 1994) and Soar (Bass, Baxter, & Ritter, 1995; Howes & Young, 1996; Miller & Laird, 1996), although each generates errors in different ways and at different levels. Fewer models exist that model error recovery, although this is clearly the next aspect to model.

A problem with models and humans is that the erroneous behavior is often task-specific; given a new task, both models and humans might not generate the same behavior. In other words, the erroneous behavior arises as a result of the combination of human, technological, and organizational (environmental) factors. Vicente (1998) delineates some of the problems in this area.

There are various taxonomies of errors that could be incorporated into models of performance. There are also other constraints that reduce the level of performance that are worth exploring, including working memory (Young & Lewis, 1999), attention, and processing speed due to expertise.

2.6 Adversarial Problem Solving

Adversarial problem solving is different from simple problem solving and makes additional requirements for modeling behavior in synthetic environments. Planning is not done within a static environment, but done in an environment with active adversaries.

Research on adversarial problem solving (e.g., Chase & Simon, 1973; de Groot 1946/1978; Gobet & Simon, 2001; Newell & Simon, 1972) has identified several aspects of cognitive behavior that have been shown to generalize to other domains, including the military domain (Charness, 1992). A key result is that players do not follow a strategy such as *minimax* but that they satisfice (Simon, 1955), that is, they satisfy themselves with a good-enough solution, which can be far from the optimal solution (de Groot & Gobet, 1996; Gobet & Simon, 1996a). This satisficing behavior can be explained by the processing and capacity limits of human cognition, such as the time to learn a new chunk or the capacity of short-term memory (Newell & Simon, 1972).

A second, related aspect is that a player's search is highly selective: only a few branches of the search tree are explored. The choice of subspace to search seems to be constrained by pattern-recognition mechanisms (Chase & Simon, 1973; Gobet, 1998; Gobet & Simon, 1996a). A consequence is that misleading perceptual cues may result in the exploration of an incorrect subspace. For example, Saariluoma (1990) reported that chess masters found a suboptimal solution when the features of the position led them to look for a standard, although inferior, subspace. The consequence for understanding combatant behavior is that pattern recognition may influence the course of action chosen as much as the detail of the way the search is carried out. In fact, de Groot (1946/1978) did not find differences in the macrostructure of search of chess players at different skill levels.

A third important result is that chess players re-investigate the same sequence of actions several times, interrupted or not by the analysis of other sets of actions. De Groot (1946) has called this phenomenon *progressive deepening*. It is related to the selective search shown by experts in other areas (Charness, 1991; Ericsson & Kintsch, 1995; Gobet & Simon, 1996a; Hoffman, 1992). De Groot and Gobet (1996) propose that progressive deepening is due both to the limits of human cognition (limited capacity of short-term memory, slow encoding time in long-term memory) and that with this searching behavior, information gathered at various points of the search may be propagated to other points, including previously visited points (this could not be done with a search behavior such as *minimax*).

These features of cognition, identified in adversarial problem solving, also occur in Rapid Decision Making (RDM) in domains such as firefighting, combat, and chess players in time-trouble. Interestingly, the model developed by Klein and his colleagues (see Klein, 1997, for a review) singles out the same features as the model developed by Chase and Simon (1973) to explain expert chess-playing: pattern recognition, selective search, and satisficing behavior.

While some aspects of adversarial problem solving are well understood, others have yet to be studied in any depth. Such aspects include the way the function used to evaluate the goodness of a state (the evaluation function) changes as a function of time, the link between the evaluation function and pattern recognition, or the learning of domain-specific heuristics, which all have direct implications for combat behavior.

Relatively little research has been done on how players take advantage of the thinking particularities of their opponent, in particular, by trying to outguess him or her. Jansen (1992) offers interesting results. He has developed a computer program that takes advantage of some features and heuristics of human cognition in simple chess endgames, such as the tendency, in human players' search, to avoid moves that lead to positions with a high-branching factor, and to prefer moves that lead to forced replies. Using these features and incorporating them in its evaluation function, the program was able to win faster (in won positions) or to avoid defeat (in lost positions) more often against human players than by using a standard alpha-beta search. In principle, such an approach could be extended to include both skill-related and individual differences in synthetic environments.

In comparison to perception and memory in games, relatively little computer modeling of human behavior has been done with adversarial problem solving (if one excludes pure Artificial Intelligence [AI] research, in which adversarial problem solving has been a favorite subject of research). One may mention the previous work of Simon and colleagues (Baylor & Simon, 1966; Newell, Shaw, & Simon, 1958), and the programs of Pitrat (1977), Wilkins (1980), and Gobet and Jansen (1994). All of these programs were created for chess and most cover only a subset of the game.

There are implications of adversarial search variation for performance (i.e., how well a planner models an opponent). This would be a natural place to model various levels of experience in opponents.

2.7 Variance in Behavior

Including more variety in how a model performs a task is one of the next steps for improving the realism of synthetic forces. Currently, many models will execute a task the same way every time and for every equivalent agent. In the real world, this is not the case. The choice of strategies and the ordering of substrategies will vary across agents and vary for a given agent across time. This lack of variance makes adversaries and allies too predictable in that they always do the same thing.

Including variance in behavior is also necessary when behavior is less predictable. Novices, with less knowledge, have greater variance in behavior (Rauterberg, 1993). In the past, variance was intentionally suppressed in simulations because it was thought that variance in real behavior was suppressed through doctrine and training. Accounting for variety in behavior is of increasing importance when modeling less-prepared and less-trained forces, and now for improving model accuracy as variance in real behavior is admitted.

Variance in behavior is also important when modeling non-combatant agents, such as white forces and civilians. These agents may be producing their behaviors deterministically, but the determiners are often hidden from other agents, making them appear relatively unpredictable. Finally, the ability to model a variety of behaviors is necessary for sensitivity analysis.

Variance will arise out of several factors. It may arise from different levels of expertise, which is covered above. It may arise from different strategies, which will require including multiple strategies and noting where orders are less likely to be followed and when panic results in orders being ignored. Variance may also arise as a type of error, such as applying a right action in the wrong circumstances.

In any case, variance in agent behavior in synthetic environments particularly needs to be included in training materials. Humans are very good pattern-recognizers—although they do not always look for or know the right pattern—and will take advantage of models that do not vary their behavior. The real opponents may not be so predictable.

2.8 Information Overload

Problems with information overload have been noted numerous times (e.g., Woods, Patterson, Roth, & Christoffersen, 1999). Hoffman and Shadbolt (1996) provide a review of work on information overload in real-time, high-workload military contexts. They also discuss challenges that information overload raises for knowledge acquisition in the context of synthetic forces environments.

Problems resolving clutter, workload bottlenecks, and finding significance in incoming data, are not yet problems for many models of human performance. Currently, most cognitive and synthetic force models do not face information overload. The situation has more typically been of a model seeing only a limited set of information and knowing how to perform only one or a few tasks.

In the near future, the models will have more complex simulated eyes as well as more knowledge to interpret the eyes' input. This will lead to more incoming information with a more difficult problem of deciding which objective to pursue next and how to choose the best strategy based on a larger set of knowledge and perceptual inputs. We will also find that models will start to have trouble with information overload, clutter, and situation assessment. Their tactics in this area will be particularly important when there are time pressures, which are common in synthetic environments and the worlds they model.