January 2004

Toward Realism in Human Performance Simulation

Barry G. Silverman
University of Pennsylvania, basil@seas.upenn.edu

Postprint version. Published in *The Science and Simulation of Human Performance*, edited by James W. Ness, Darren R. Ritzer, & Victoria Tepe (New York: Elsevier, (2004), pages 469-498.

Toward Realism in Human Performance Simulation

Abstract
This chapter focuses on challenges to improving the realism of socially intelligent agents and attempts to reflect the state of the art in human behavior modeling with particular attention to the impact of values, emotion, and physiology/stress upon individual and group decision-making. The goal is to help those interested in constructing more realistic software agents for use in human performance simulations in both training and analysis settings. The first two sections offer an assessment of the state of the practice and of the need to make better use of human performance moderator functions (PMFs) published in the behavioral literature. The third section pursues this goal by providing an illustrative framework for integrating existing PMF theories and models, such as those on physiology and stress, cognitive and emotive processes, individual differences, and group and crowd behavior, among others. The fourth section presents asymmetric warfare and civil unrest case studies to examine some of the concerns affecting implementation of PMFs such as verification, validation, and interoperability with existing simulators, artificial life emulators, and artificial intelligence components. The final section of this chapter concludes with lessons learned and with some challenges if the field is to reach a greater level of maturity.

Comments
Postprint version. Published in *The Science and Simulation of Human Performance*, edited by James W. Ness, Darren R. Ritzer, & Victoria Tepe (New York: Elsevier, (2004), pages 469-498.

Chapter 9
Towards Realism in Human Performance Simulation

Barry G. Silverman, Ph.D.
Ackoff Center for Advancement of Systems Approaches (ACASA), Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104-6315, USA. e-mail: barryg@seas.upenn.edu

ABSTRACT

This chapter focuses on challenges to improving the realism of socially intelligent agents and attempts to reflect the state of the art in human behavior modeling with particular attention to the impact of values, emotion, and physiology/stress upon individual and group decision-making. The goal is to help those interested in constructing more realistic software agents for use in human performance simulations in both training and analysis settings. The first two sections offer an assessment of the state of the practice and of the need to make better use of human performance moderator functions (PMFs) published in the behavioral literature. The third section pursues this goal by providing an illustrative framework for integrating existing PMF theories and models, such as those on physiology and stress, cognitive and emotive processes, individual differences, and group and crowd behavior, among others. The fourth section presents asymmetric warfare and civil unrest case studies to examine some of the concerns affecting implementation of PMFs such as verification, validation, and interoperability with existing simulators, artificial life emulators, and artificial intelligence components. The final section of this chapter concludes with lessons learned and with some challenges if the field is to reach a greater level of maturity.

1. Introduction

The fields of virtual reality and microworld simulation have advanced significantly in the past decade. Today, computer generated personas or agents that populate these worlds and interact with human operators are now used in many endeavors and avenues of investigation. A few of many example application areas are Hollywood animations for movies, cartoons, and advertising (von-Neuman & Morganstern, 1947); immersive industrial and safety training simulations (Fudenberg & Tirole, 2000; Silverman et al., 2001); distributed, interactive military war games and mission rehearsals (Johns & Silverman, 2001); and personal assistant agents to reduce technologic complexity for the general public, among others (Weaver, Silverman, Shin, and Dubois, 2001).

A common challenge running throughout these applications is to increase the realism of the synthetic agents' behavior and cognition. This is not an idle fancy, but a serious objective that directly affects the bottom line of commercial concerns, mission achievement in non-commercial organizations, and the safety and health of individuals who need to transfer skill sets from virtual to real worlds. Agent-oriented products that are more emotively natural and offer a better cognitive fit tend to sell better, such as the successful games *Tamagotchi* or *Catz and Dogz* (El-Nasr, Ioerger, and Yen, 1999). This lesson applies to embedded agents as well as stand-alone products. People are known to anthropomorphize technologic items such as cars, slot machines, computers, ATM machines, etc. A strategy of beating the competition is beginning to emerge by including greater degrees of personality, human modes of interactivity (e.g., voice synthesis for car navigation systems), and emotive features in personas embedded ubiquitously (e.g., lip-synched and facially-accurate expressions) (e.g., see Nass, 1996; Hayes-Roth, 1998; Wayner, 1995). Similarly, in training, analytical, and education systems with military applications there is a growing realization that greater cognitive subtlety and behavioral sensitivity in synthetic agents provides human trainees with both (1) more and better opportunities to explore alternative strategies and tactics, and (2) higher levels of skill attainment (e.g., see Downes-Martin, 1995; Sloman & Logan, 1999; Angus & Heslegrave, 1985). These benefits are possible if the tactics, performance, and behavior of synthetic agents changes in response to alterations in an array of behavioral and cognitive variables. As a few examples, one would like agent behavior to realistically change as a function of their assigned native culture (vital for executing missions against forces from different countries), their level of fatigue and stress over time and in different situations, and/or their effectiveness following the loss of an opposing force's leader.

1.1 Challenges for Increasing the Realism of Human Behavior Models

There has been significant progress toward the development of improved human-like performance of synthetic agents. However, several serious problems continue to challenge researchers and developers.

Developers have insufficient behavioral knowledge. To date, models of emotivity and behavior that have been commercialized still tend to be shallow and unsatisfying. There is no deep model of human-agent interactivity. Synthetic agent forces are naïve and unable to act with the credibility or behavioral variety seen in human operators.

Artificial life has focused on low level cognitive functionality. Character animators have created virtual life forms (e.g., fish, plants, talking heads, full body characters, and groups) that are physically realistic, geometrically accurate, and kinesthetically natural when moving about within their virtual settings (e.g., see Badler, Chi, and Chopra, 1999; Badler, Palmer, and Bindiganavale, 1999). There has even been significant development of architectures to produce animated characters that react appropriately to a small range of emotive and environmental stimuli such as fright and flight, flocking, and lip- and facial-movement-synching to utterances or stimuli: (e.g., EBAA, 1999; Terzopoulos, 1999). However, these tend to be reactive systems that perform no deliberative or high-level decision making or cognitive processing such as has been conceived by the artificial intelligence community (e.g., see Funge, 1998; Rickel & Johnson, 1998; Tambe et al., 1995).

Artificial intelligence (AI) focuses on high level cognitive functionality. Researchers in the "rational agent" community have created a wide array of methods, often formal and grounded logics, to support agent reasoning (Bradshaw et al., 1999), inter-agent communications (Labrou, Finin, and Peng, 1999), and autonomous planning and learning (Laird et al., 1995). These methods make it possible for unembodied agents to sense and respond to their virtual environments. However, extensive computing resources are necessary to support such abilities. It remains to be seen whether the necessary computing cycles can also be designed to support similar realistic abilities in relatively more resource-intensive embodied life characters. In addition, care must be taken when applying artificial intelligence to enhance the behavioral and cognitive fidelity of synthetic characters. It is easy to use these techniques to create capabilities that no real human being would possess. For example, a rational agent can perform its tasks without suffering the effects of fatigue, stress, heat, illness, biochemical exposure, or other factors that would likely affect the performance of a real human operator. Surprisingly, this seems to be true of widely respected 'cognitive models' (e.g., SOAR, Laird et al., 1995; ACT-R, Anderson, 1990) whose AI-based designs tend to ignore much that is known about how cognition varies as a function of individual differences, situational change, and task attributes.

Behavioral and cognitive researchers tend to ignore implementation. There are well over one million pages of peer-reviewed, published studies on human behavior and performance as a function of demographics, personality differences, cognitive style, situational and emotive variables, task elements, group and organizational dynamics, and culture. This is a potentially rich resource for agent developers. Unfortunately, almost none of the existing literature addresses how to interpret and translate reported findings as principles and methods suitable for implementation or synthetic agent development (Silverman, 1991). Too often, factors described in the human performance literature are only roughly quantified. Informed judgment and/or additional testing is required to parameterize factors as dose-response curves or PMFs. It is time consuming and sometimes beyond the abilities of laypersons (agent builders) to determine the validity and generalizability of findings reported in behavioral scientific studies.

There is a dearth of interchange standards. There are few interchange standards in the AI and game-maker communities. One cannot readily substitute characters or agents from one simulation or game to run in the world of another. Existing systems cannot easily be modified or extended to incorporate interesting new factors, findings, or code. Interchange standards are needed to exploit the rich diversity of achievements from various related fields of development. At a basic level, such standards would cover API specifications for plug and play modules. Far more complex standards could be developed to support the exchange of agent knowledge and ontologies, behavior models, and the means to apply them within a diverse variety of systems.

There is a need to validate "useful" models of human behavior. According to folkism, "all models are broken, some are useful." No model will ever capture all the nuances of human emotion, the full range of stress effects, or how these factors affect judgment and decision making. However, to the extent that a model provides a valid representation of human behavior, it will likely be useful to those who wish to simulate that behavior. Unfortunately, the problem of model validity has no simple solution. It is difficult to run existing simulations against real past events. Many first principle models from the behavioral science literature have been derived within a particular setting, whereas simulation developers may wish to deploy those models in different contexts.

Likewise, there are validity issues raised by AI models of agent reasoning that are able to simulate human task performance in the absence of human-like reasoning. Is it even possible to validate the integration of multiple factors (e.g., stress and fatigue) when the research literature is largely limited to the study of independent rather than interactive effects of these factors? What methods of validation should be used, and for what purposes are models to be considered trustworthy? These are just a few of the many issues and questions that must be resolved in order to identify whether and how an agent-based model of human behavior should be implemented.

2. Human Behavior Modeling for Virtual Agents

To overcome the obstacles identified above, it would be useful to remove cross-community barriers and create a set of sharable resources for the modeling and simulation community. This goal is considered here, beginning with a discussion of what can be learned from the behavioral science literature and then moving on to the question of how those assets can be integrated with existing, 'reusable' models of human performance and cognition.

There is a voluminous literature, easily numbering in the 10,000s of studies, on the subject of human performance under stress. One of the earliest studies in this field generated the now classic Yerkes-Dodson "inverted u" curve, which demonstrates that as a stimulus or moderator is increased, performance is initially poor, then improves, and then falls off again after passing a threshold (Teigen, 1994). Thus, performance can be better in slightly chaotic, moderately time-pressured settings than in settings absent of stress. Some of the best reviews of this literature can be found in military meta-analyses (e.g., Driskell et al., 1991) and other information resources such as The Engineering Data Compendium, which includes many classic studies and useful surveys on specific PMFs (Boff & Lincoln, 1988). The Virtual Naval Hospital (www.vnh.org) addresses the many dozens of factors that may effect combat stress.

In the field of medicine, a similarly vast amount of data, findings, and lessons learned are generated by clinical trials that revolutionize medical practice. Because each clinical trial has its own unique strengths and design flaws, a voluntary international effort has been launched to share result sets (raw data, study designs, and conclusions) in evidence-based repositories that include a "structured abstract" for each study. These abstracts are written by qualified reviewers who attempt to extract each study's highlights and guidance, and to provide a validity assessment of the utility and applicability of the results. This approach has become popular. Many volunteer reading groups and journal clubs now routinely author structured abstracts and many medical journals now require that clinical trials include structured abstracts as a condition of publication.

A comparable approach is possible in the field of human performance modeling. The goal in this case would be to identify performance moderator functions (PMFs) and related human behavior models from within the research literature and (1) identify and properly abstract them, (2) assess their internal validity, and (3) prepare the best of them for implementation and reuse. This would make it easier to (re)utilize PMFs and thus to improve the realism of human behavior in models and simulations.

(b) Validity Assessment Scale

- **Title:**
- **Authors:**
- **Organization:**
- **Reference Number:** **Date:** **Pages:**

  - <u>**TASK**</u>
- **Domain:**
- **Echelon:**
- **Tasks Studied:**
- **Cognitive Framework:**

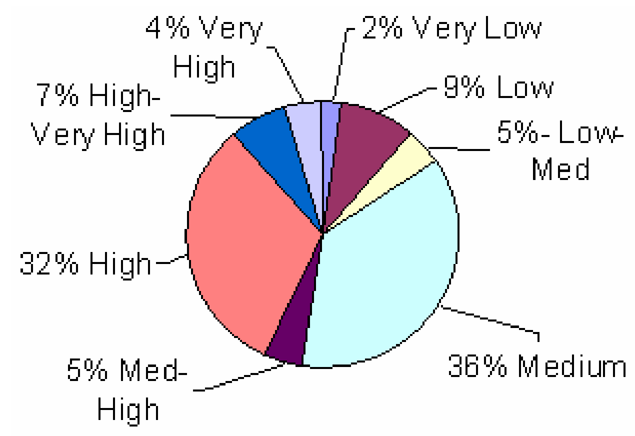
  - <u>**METHODOLOGY**</u>
- **Study Goal:**
- **Study Procedure:**
- **Number of Subjects:**
- **Arms of Study:**

  - <u>**FINDINGS**</u>
- **Performance Moderator Variables:**
- **Performance Moderator Functions (PMF):**
- **Modeling Technique:**
- **Lessons Learned:**
- **PMF Validity Info:**

<u>**ORIGINAL ABSTRACT**</u>

<u>**CONCEPTUAL MODEL FRAMEWORK (CMF)**</u>

Scale:	Degree of Value of Literature Item for Constructing PMFs
5= VERY HIGH	PMFs provided with backup data sets
4= HIGH	Could make PMFs directly from the data in this study
3= MEDIUM	Some preliminary data for initial PMF construction, but more data needed
2= LOW	Theoretical model suggested from which an ungrounded PMF could be derived.
1= VERY LOW	No valid data for PMF construction
0= NONE	Irrelevant to the PMF construction process.



<u>**(c) Percent of PMFs in Each Validity Category (based on a sample of 486 PMFs in the Anthology)**</u>

Figure 1. Validity of the collection of PMFs in the anthology

Figure 1a shows the template of a structured abstract containing several sections. The top of the template includes a Reference section, which is largely useful for indexing and search purposes. Likewise, the Task section serves a vital role in helping to organize the abstract within various taxonomic categories useful in the search for collections. The lower half of the template in Figure 1a focuses on issues related to evaluation. These sections are designed to help a reader quickly determine the study's PMFs (Findings section) as well as the study's design strengths and weaknesses (Methodology section). The Findings section includes a field on the study's validity and lessons learned. The Methodology section tells the reader how the PMFs were derived, what types of subjects were used in the study, and what conditions were evaluated. Finally, the template includes the study's original abstract and a section on the Conceptual Model Framework (CMF) which includes a mapping of the study's PMFs within a common mathematical framework (see Section 3, this chapter).

To assess validity for PMF construction, Figure 1b proposes a 5-point rating scale ranging from 5 for VERY HIGH validity to 1 for VERY LOW validity, and a sixth category (0) reserved for cases in which the study is judged to be irrelevant. By this approach, validity increases to the extent that a study is grounded in empirical data (vs. theory alone) and to the extent that it contains PMFs that can be utilized by the modeling and simulation community.

It is commonly believed that the existing behavioral science literature offers little in the way of results directly valuable to the extraction of PMFs for use in Human Behavior Models (HBMs). Pew and Mavor (1998, p.242) refer to the "individual difference variables about which the least is known so that they cannot at this point be encoded directly into a model."

As Figure 1c shows, part of this negative prediction is born out. Based on a sample of 486 PMFs in an HBM anthology we created (see Silverman et al., 2001), only 4 percent of studies had VERY HIGH validity, offering empirically-grounded PMFs directly available for use with no additional manipulation. However, about 30 percent of the studies from this sample had HIGH validity – that is, they were well-grounded in terms of their design and data availability, and offered immediately useful data for constructing PMFs. More encouraging still is that an additional 34 percent of the sample studies could be turned into temporary working draft PMFs. Nineteen percent of the reports surveyed presented theoretical frameworks of interest and value to the development of human behavior models. Only 19 percent of the studies were judged to be entirely useless (VERY LOW validity, or NONE).

3. Integrating the Many PMFs into a Cognitive Framework

The PMF findings discussed above suggest that there is a large, untapped resource to assist those who create human performance simulations. Models from the behavioral science literature have been ignored by the various agent modeling communities for a variety of reasons. However, much can be gained the if simulation developers begin to embrace such collections and work to implement and test their results. This, in turn, should and could create a forcing function back upon behavioral scientists. Behavioral science has produced some compelling models and PMFs of individual factors, but they still need to do a better job of studying and evaluating integrative frameworks.

This section will explore what one such integrative framework might look like. There are a large number of similar frameworks in the literature (e.g., a useful comparison of 60 such models may be found in Crumley & Sherman, 1990). The framework described here, known as PMFserv (Silverman et al., 2001; Silverman, Johns, O'Brien, Weaver, and Cornwell, 2002; Silverman, Johns, Weaver, O'Brien, & Silverman, 2002), is not intended as the best cognitive architecture or agent algorithm but rather as a reasonable framework within which many contributions from the literature could be integrated, investigated, and extended as needed. This framework does not replace existing PMFs, existing AI/cognitive models, or existing artificial life approaches. Instead, this framework attempts to show how one needs *all* of these approaches, and others, in order to produce a realistic human performance simulation system.

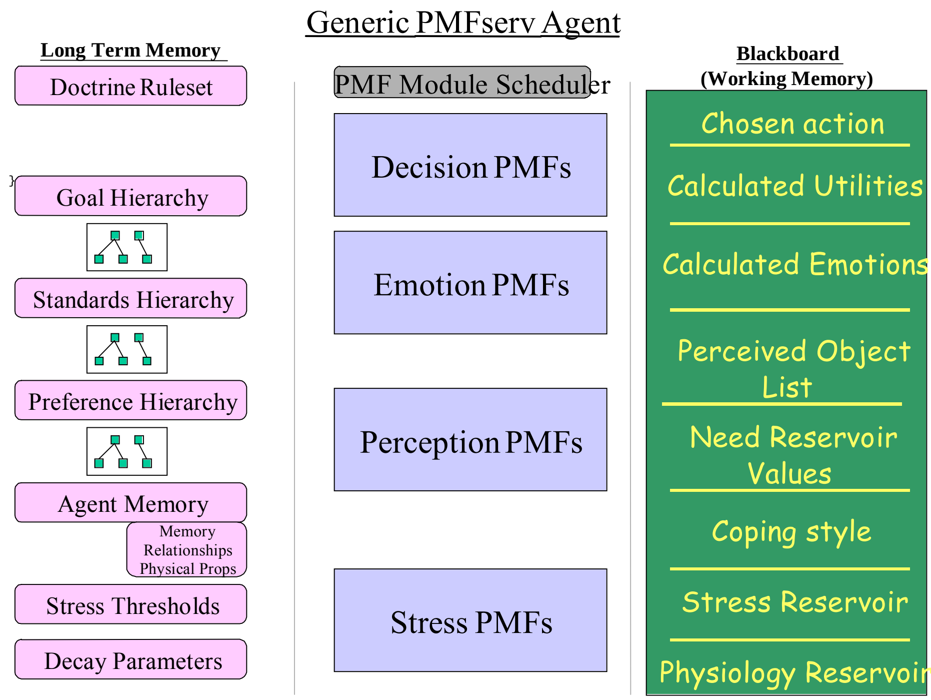


Figure 2. PMFserv overview

The principal feature of PMFserv is that it models human decision-making based on emotional subjective utility constrained by stress and physiology. As such, PMFserv is an agent architecture in its own right, with the flexibility to act as a meta-level emotional arbitrator for others' cognitive architectures or to provide a fully functional stand-alone system to simulate human decision making. PMFserv is built around a 'blackboard' data structure that loosely corresponds to a short-term or working memory system. Sensory data about the world flows into the lower layers of the blackboard structure, as constrained by stress and other factors described below. Modular PMF subsystems then manipulate data contained in the blackboard and in a long-term memory store. Information is layered on the blackboard such that each layer is dependent on the layers below it for any given agent decision cycle (see Figure 2).

Moving up the blackboard from the bottom reveals the decision cycle of a single agent. Physiological data across a range of measures (including PMFs for arousal, exertion, hunger, thirst, injury, etc.) are combined to set the levels of a series of stress reservoirs, as in Figure 3. Each reservoir keeps track of both the current level of the stimulus in the environment and any stress that results from that stimulus. There are a large number of stressors that moderate an agent's ability to perform up to capacity. In some cases, these produce alarms. For example, alarms may occur when there is pain or when a critical threshold is exceeded (e.g., hunger, fatigue, panic, etc.). An important criterion for such a module is that it should support study of common questions about performance moderators (e.g., easy addition or deletion of reservoirs such as pain or stress), individual differences in reacting to particular stressors, and/or how to model reservoir behaviors linearly (PMFserv approach) or non-linearly, such as with bio-rhythms.

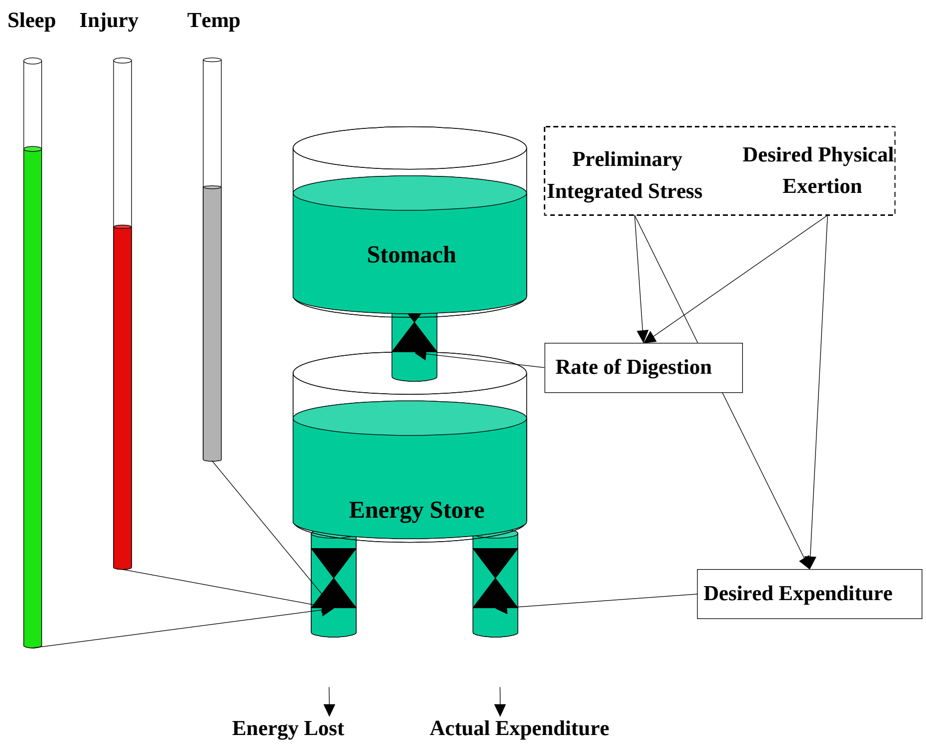


Figure 3. Physiology module uses PMF reservoirs

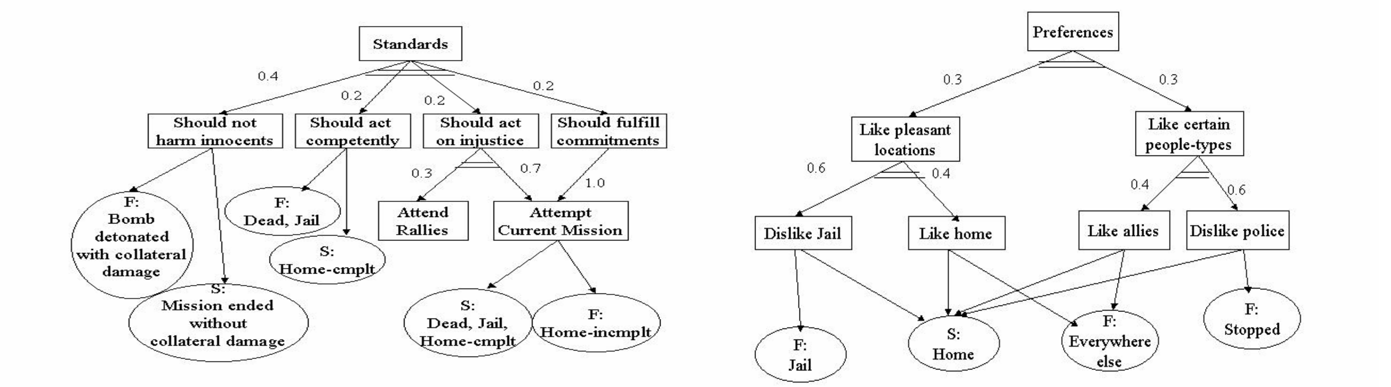
The PMF literature is useful for modeling an individual's stress factors. However, a major gap in the literature is that very few studies have been conducted to determine how multiple factors combine to produce overall or integrated stress in an individual (e.g., Hammond, 2000). One approach is that of Hendy and Farrell (1997), who adopt an information processing theory and model that focuses on equipment and screen operators and includes factors such as task workload (bits to process) and work rate (bits/sec). They offer an information processing (bit throughput, error rate, decision time) account that attempts to explain the effects of time pressure, task difficulty, mental capacity, fatigue, motivation, anxiety, and the like. However, they offer little data to support their model.

Hursh & McNally (1993) reviewed 1,300 studies to develop a model of decision making in battle that focuses solely on effectiveness under stress. Gillis and Hursh (1999) later extended this model to account for what they claimed were the prime determinants of (stressed or non-stressed) performance: effective fatigue (summarized over PMFs), event stress, and time pressure. We found this to be a reasonable solution until more rigorous models are derived and defended. PMFserv thus tracks these three stress 'reservoirs' and also combines them heuristically to compute an overall Integrated Stress (Ω) estimate.

An integrated stress estimate provides a useful basis for use of Janis and Mann's "Conflict Decision Theory," which has been derived from years of analyses of many subjects under stress. Conflict Decision Theory is robust and its validity has been supported through a meta-analysis of the literature (Janis & Mann, 1977). In this model, there are five coping modes, all but the third of which bound an agent's ability to fully perceive its environment and make rational decisions based on those perceptions. In mode 1 (Unconflicted Adherence), the agent does not update its perceptions about the world and continues doing whatever it was doing during the preceding tick of the clock. In mode 2 (Unconflicted Change), the agent does not update its perceptions about the world, but uses those outdated perceptions to formulate its present course of action. In mode 3 (Vigilant), the agent updates its perceptions and reaches a decision based on which action will be most useful. In mode 4 (Defensive Avoidance), the agent updates some of its perceptions, but fails to update its perceptions concerning those objects that cause the most negative event stress. In mode 5 (Panic), the agent either cowers in place or flees, depending on the average value of its emotions from step 2. PMFserv uses its calculated integrated stress value (Ω) to determine the agent's coping mode in each decision cycle. The stress thresholds at which agents shift between coping modes can be set on a per-agent basis, thus allowing for individual differences in reaction to stress, which in turn affects individual decision-making ability.

Before describing an agent's decision making and how coping modes (and integrated stress) serve to constrain decision making, it is first useful to understand two intertwined modules on the PMFserv blackboard. These modules are emotion and perception. According to current theories (Damasio, 1994; Ortony, Clore, and Collins, 1988; Lazarus, 1991), the emotion module receives stimuli from the perception module (see below) as moderated by the physiological system. It includes long-term memory as a set of values (modeled as trees) activated by situational stimuli as well as any internally-recalled stimuli. These stimuli and their effects act as releasers of alternative emotional construals and intensity levels. Emotional activations in turn provide somatic markers that assist the agent in recognizing problems, potential decisions, and actions. In order to support research on alternative emotional construal theories, this subsystem must include an easily alterable set of activation/decay equations and parameters for a variable number of emotions. Further, since construals are based on value trees, this module must serve as a value tree processor and editor. Simply by authoring alternative value trees, one should be able to capture the behavior of alternative "types" of people and organizations and predict how differently they might assess the same events, actions, and artifacts in the world around them. This requires that the emotion module be able to derive the elements of utility and payoff that the decision module will need to make choices.

PMFserv's emotion unit uses a commonly implemented model called OCC (abbreviation in homage to psychologists Ortony, Clure, and Collins, 1988). The general idea is that an agent possesses *Desires or Goals for action*; *Standards for behavior of self and others*; and *Preferences for people, objects, and situations*. PMFserv models these motivators as multi-attribute value trees called GSP Trees (Figure 4). An action in the simulated world can be represented by a series of successes and failures on the sub-nodes of these three trees. Each child node on a tree is given a weight that describes how much it contributes



4a -- Terrorist's Standards for Agent Behavior/Action		4b -- Terrorist's Preferences for Artifacts or Objects

S = Succed, F = Fail 

Figure 4. Illustrative shred of GSP trees and values of a sample terrorist

to its parent node. To determine the emotional utility of an action or event, PMFserv multiplies the degree of success and failure of each node up the trees. From the top nodes on each tree, 11 pairs of oppositely valenced emotions are generated. A few examples of these are:

Joy amount of success on the agent's top goals node
Distress amount of failure on the agent's top goals node
Pride amount of success on the agent's top standards node
Shame amount of failure on the agent's top standards node
Liking amount of success on the agent's top preferences node
Disliking amount of failure on the agent's top preferences node

PMFserv allows for a common set of *Goals*, *Standards*, and *Preferences* trees whose structure is shared by all agents. However, the tree weights are unique for each agent and thus capture individual differences that may be determined by culture, ideology, or personality. When these trees are applied to the task of selecting a next action, they give each agent a robust and individual worldview. When they are applied to immediate simulated events, emotions derived from the top of the *Goals* tree provide an estimate of the individual agent's event stress, as mentioned earlier.

When contemplating a next action to take, the agent calculates the emotions it expects to derive from every action available to it, as constrained by perception and coping style. We assume that utilities for next actions, a_k, are derived from the emotional activations. Silverman, Johns, Weaver et al. (2002) describe the set of mathematical equations for the use of the OCC model to help generate up to 11 pairs of emotions with intensities (I_ξ) for a given action. These expressions capture the major dimensions of concern in any emotional construal – values, relationships, and temporal aspects. Utility may be thought of as the simple summation of all positive and negative emotions for an action leading to a state. Since there will be 11 pairs of oppositely valenced emotions in the OCC model, we normalize the sum as follows so that utility varies between –1 and +1:

$$U = \sum_{\xi} I_{\xi}(a_k)/11 \qquad [1.0]$$

While one can argue against the idea of aggregating individual emotions, this summation is consistent with the somatic marker theory. One learns a single impression or feeling about each state and about actions that might bring about or avoid those states. The utility term, in turn, is derived dynamically during each iteration from an emotional construal of the utility of each afforded action strategy relative to that agent's importance-weighted value ontologies (GSP trees) minus the cost of carrying out that strategy.

For this to work, the agent must use its perception module, as constrained by coping mode and emotive needs, to see what's going on in the world. Perception should be focused based on an agent's physiology, coping style, prior emotional needs, and any memory elements that might have been created before the current cycle. For example, if the agent's coping mode is Panic or Unconflicted Adherence, it will not notice anything new in the world. Otherwise, PMFserv applies affordance theory (Gibson, 1979) such that each object in the simulated world executes perception rules to determine how it should be perceived by the agent and generates a list of the corresponding actions (a_k) and affordances it can offer that agent (e.g., a rock indicates it can be thrown, which will afford success in hurting an opponent and will consume *x* units of energy). These affordances provide reservoir replenishment or drawdown impacts and GSP tree multipliers for degree of leaf node success or failure. In this fashion, PMFserv agents implement situated ecological psychology (Gibson, 1979).

The decision module serves as the point where diverse emotions, stressors, coping style, memories, and object affordances are all integrated into a decision for action (or inaction) to transition to a new state (or remain in the same state). In essence, at each tick of the simulator's clock, each agent must be able to process the following information: current state name (or ID); stress-based coping mode (Ω_i where $i = 1,5$); currently afforded transitions and what action might cause those state transitions (a_{nm} in $A(\Omega)$); and subjective desires for each state based on 11 pairs of emotional scales summed into an overall utility score, U. Using all of this information, the agent must select a decision style (Φ, defined below) and process the information to produce a best response (BR) that maximizes expected, discounted rewards or utilities in the current iteration of the world. The decision module is thus governed by the following equation:

BEST REPLY (BR_t) = $\Phi_{iSTRESS,\,\Omega}\{u_{mn}(s_t, a_{mnt}), p_{mn}\}$, subject to $a_{mnt} \in A(\Omega)$ [2.0]

Where,

$\Phi_{iSTRESS,\,\Omega}\{.\}$ = as defined below for the alternative values of Ω

p_{mn} = perceived probability = $(1 - \Delta)\, e_m + \Delta_{m\tau}\, p_{m\tau}$

u_{mn} = $(1-\delta)$ x (U from equation 1.0)

Δ = memory coefficient (discounting the past)

τ = number periods to look back

$$e_m = \begin{cases} 0 & \text{action m not situationally relevant} \\ 1.0 & \text{action m is situationally relevant} \end{cases}$$

δ = expectation coefficient (discounting the future)

$A(\Omega)$ = action set available after coping mode-constrained perception

This is nothing more than a stress-constrained subjective-expected utility formulation. Utility may be thought of as the simple summation of all positive and negative emotions for an action leading to a state. While one can argue against the idea of aggregating individual emotions, this summation is consistent with Damasio's somatic marker theory (Damasio, 1994). One learns a single impression or feeling about each state and about actions that might bring about or avoid those states. Also, there is a large literature on decision style functions (e.g., among many others see Tambe et al., 1995; Bradshaw et al., 1999; EBAA, 1999; Terzopoulos, 1999; and Funge, 1998), and the decision processing style function, Φ_Ω, merely indicates that there is a rich set of possibilities that one can explore within the framework proposed here. Thus, in Vigilant mode one might invoke SOAR, ACT-R, COGNET, or others. Alternatively, simulated experts can adopt the Recognition Primed Decision Making (Klein, Orasanu, Calderwood, and Zsambok, 1993) style, while novices will tend to use a more traditional decision tree.

The algorithm proposed above applies Conflict Theory where appropriate. That is, if the agent's coping mode is Panic or Unconflicted Adherence, no alternatives are weighed and the agent will execute its panic behavior or continue to do what it had already decided to do in the last cycle. Likewise, Unconflicted Change prevents any planning, and the agent must follow the next step of any existing plan. Only when stress increases and the agent's coping mode shifts to Vigilance can the agent re-plan (with any Φ_Ω method as desired).

4. Making PMFs Useful

This chapter began by lamenting that many useful contributions from the behavioral science literature aren't being used to improve existing simulations, artificial life, and artificial intelligence systems. The previous section presented a framework for bridging that gap. To achieve full integration, a reasonable framework is necessary but not sufficient. There are a number of additional issues that must also be addressed in order to achieve meaningful implementation of PMFs. This section will provide an overview of these issues and address possible approaches to dealing with them.

4.1 Description versus Prediction

Pew and Mavor (1998, p. 268) point out that the short-term research goals for the modeling community should be to "apply existing knowledge about both extrinsic and internal behavior moderators to establish value settings for various parameters of human behavior … and observe the effects of the use of such estimates in a sample of simulated engagements." Until this type of work is undertaken, it will be very difficult for the modeling community to utilize the literature on behavior moderators.

For these and other reasons, it is essential to scientifically investigate and more thoroughly document the properties of PMFs. The most reasonable way to do this is to observe them in use. That is, one would like to postulate various forms for the PMFs and study how reasonable those forms are, what impact they have on agent reasoning, how they combine dynamically and under stochastic conditions, and how sensitive are reasoning performance, workload, and outcome effects to small changes in the shape of PMFs and in the size of various weights.

The first concern thus pertains to what one expects from a simulation. In general, human performance simulations are used to (1) analyze strategies and plans, doctrine and tactics, work efficacy studies, and material design and acquisition choices; or (2) train personnel for individual combat-related skills, for leadership and coordination capabilities, and for mission rehearsals. Many analysts hope that human performance simulations will predict the future or its potential outcomes. Given the relative newness of human behavior modeling as a science, this is not an appropriate expectation. One should ideally try to use a human performance simulation to explore the space of analytic possibilities or to provide a range of reasonable situations for training. For example, when using a framework such as PMFserv to specify a scenario for analysis or training, one begins by designating each type of character in the scenario (e.g., green recruit, seasoned combatant, or worn out veteran) according to that character's default reservoir rates and thresholds, coping mode cutoff points, GSP tree values, and decision style options. These are all mean settings, however. One can then run the simulation in Monte Carlo style wherein each rate, threshold, or value is perturbed via a random number seed around the mean so as to sample across the distributional space. Analysts are generally comfortable with Monte Carlo simulations and with making the numerous runs of the simulator in order to adequately describe the space of possible outcomes. However, training developers often find it necessary to treat all trainees alike and so will tend to fix on a given or "interesting" seed of the random number generator and train to that (hopefully most challenging) outcome set. Only by having trainees repeat the training will they begin to see and appreciate the range of possible outcomes that can occur and learn how to prepare for those many possibilities.

4.2 Verification Testing

A second concern is how to reliably determine whether each agent is operating according to specification. That is, verification is necessary to ascertain that agent behavior is (1) consistent with respect to individual PMFs; (2) complete with respect to the collected set of all PMFs being implemented; and (3) somehow coherent with respect to their own goals, standards, and preferences in the scenario. In an effort to better understand how

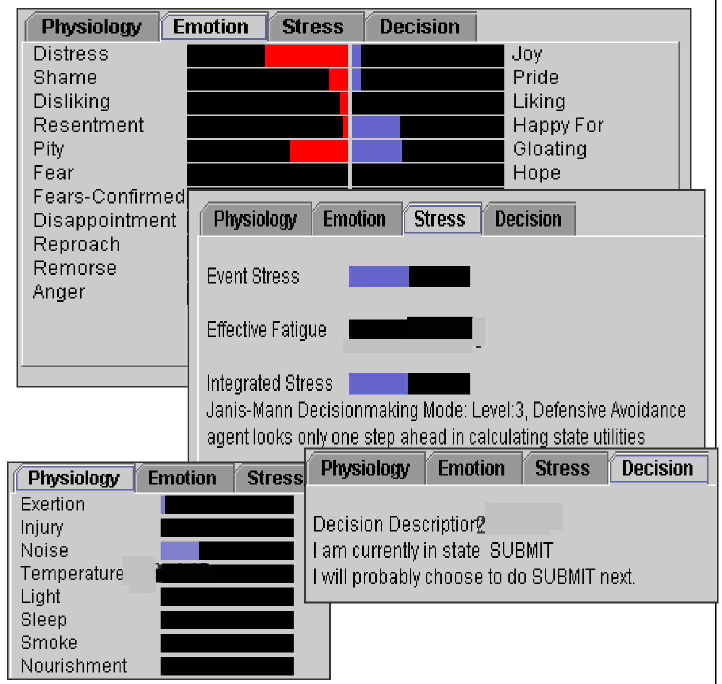
<table>
<tr>
<th>a) Defenders(D) at a Bridge Checkpoint as Civilians (N) and Terrorist Agents (A) Deboard a Yellow Schoolbus</th>
<th>b) Visual Interface to Each Agent’s PMFs (Neutral female shield in Coping Mode: Defensive Avoidance, Emotions: Mixed, State: SUBMIT to being a shield)</th>
</tr>
<tr>
<td>
</td>
<td>
</td>
</tr>
</table>

Figure 5. An implementation using General Gorman’s Fombler’s Ford scenario

PMFs operate, a number of demonstration scenarios were built using PMFserv. One of the earliest demonstrations tested was a simulated ambush at a checkpoint inspired by a similar scenario depicted in GEN Paul Gorman’s, *In Defense of Fombler’s Ford* (Gorman, 2000). This situation involved a school bus deboarding women and children (“neutrals” or N) with six terrorists lurking in their midst (“attackers,” A). As Figure 5 shows, the group of passengers deboards near a bridge checkpoint where several “defenders” (D) are unaware of the ambush. PMFserv manages the physiology, stress, emotions, and decisions of each of the agents in this scenario, permitting individual agent reactions to emerge bottom up as they interact with the scene and with events that unfold. The defenders’ standards (including orders) make it difficult for them to shoot civilians, while the attackers’ standards permit them to treat civilians as shields because the attackers have a goal to take bold action and harm their opponents. To that end, the attackers’ regard civilians as “objects” to be manipulated.

One way to verify that the PMFs are working properly is to separately examine each agent and each PMF as the scenario unfolds. In the example above, we authored a set of visual PMF interfaces that may be examined by double clicking on any given agent. Figure 4b depicts these visuals for one of the female shield agents. Its various tabs reveal her current physiology (mild exertion, noise), coping mode (Defensive Avoidance), emotions (e.g., liking or disliking specific aspects of the situation, pitying those having to act as shields but gloating over the impending defenders' fate), and her decision to submit to being a shield. These displays allowed us to see which PMFs are working or broken. After a debugging interval when we thought all PMFs were working properly, the scenario still failed to unfold properly. That is, no one died, despite the fact that all agents converged at the bridge and all attackers and defenders discharged their weapons continuously (this is represented by the black lines emanating from some of the agents). Upon closer inspection of the PMF set, we noticed that everyone's noise PMF was relatively elevated. This was causing a fair amount of arousal and all agents were converging at the bridge in an emotionally elevated state. We then looked at possible causative factors and discovered that the weapon aiming routine had not been calibrated. Thus, all agents were shooting vertically up in the air. Once we corrected this, verification was complete, and the simulation was able to produce various outcomes depending on the fatigue or alertness of the checkpoint defenders.

4.3 Validation via Correspondence Testing

Verification that multiple PMFs work in concert is not the same as validation. The latter requires one to evaluate how well scenario outcomes correspond to real world or historical events. Historical recreations are challenging because participants' thoughts, motivations, and stress levels can be known or estimated only at a general level. There are different ways to approach this problem. As a qualitative approach, one might ask knowledgeable observers to compare the simulated and historical outcomes. A more quantitative approach would be to quantify events along a timeline and/or quantify outcomes by type of participant and determine correlative relationships between real and simulated events and outcomes. Of course, it is also possible to combine qualitative and quantitative efforts to evaluate correspondence.

PMFserv has not yet been fully studied for its quantitative correspondence to real world scenarios. However, it has been tested against scenarios that depict civil disturbances. Each of these scenarios featured a crowd that had gathered to protest a perceived social injustice. In one series of scenarios, the crowd protested a roadblock that prevented people from going to work. In other scenarios, crowds protested outside a prison. All of these scenarios featured similar characters. The group of protesters included men and women, employed and unemployed. Each scenario also included police officers, onlookers, and one or two instigators who tried to rouse the crowd. No outcome was programmed into the simulation's rules or equations. However, significant effort was expended to develop and model appropriate GSP trees (see previous section) and personal value weights for each agent. Individual agents then made their own (micro)decisions that led to emergent macro-behavior.

In the various scenarios tested, we evaluated the impact of diverse PMFs (e.g., alternative personal and cultural value levels, impact of chanting and taunting, and diverse security doctrine/orders and behavior) on crowd behavior and on when new crowd equilibria emerged (e.g., peaceful protest, scatter, riot). These efforts enabled us to document a number of lessons learned about the replication of anticipated emergence of different types of crowd behavior (Silverman, Johns, O'Brien, Weaver, and Cornwell, 2002;

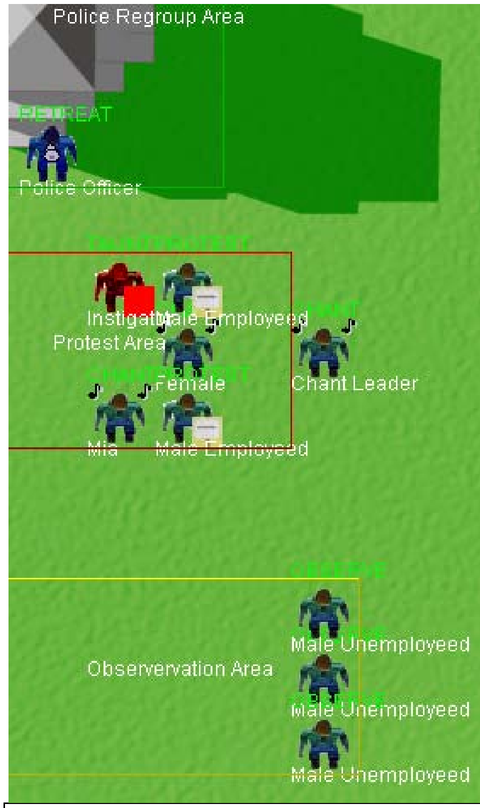
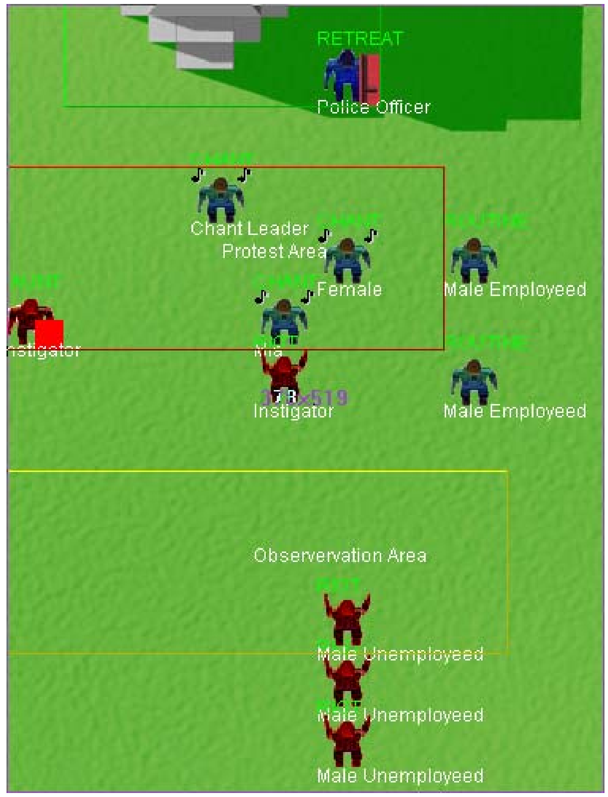
(a) Organized Protest Mode With Instigator and Too Few Security	**(b) Unemployed Males Moving to Loot, Rest of Crowd About to Flee**
	

Figure 6. The crowd scenes seem to correspond to many real world events

Silverman, Johns, Weaver, O'Brien, and Silverman, 2002; Silverman, 2001; Cornwell, Silverman, O'Brien, and Johns, 2002; Johns & Silverman, 2001). As an example of correspondence checking, the crowd literature (Horowitz, 2001; McPhail & Wohlstein, 1983) indicates that looting tends to occur when young unemployed males (who rarely join organized activities) take advantage of chaos and distracted security forces. In our simulations, female protesters and employed men tended to flee from riot situations, while unemployed men lurked on the fringes of the protest scene and then proceeded to loot (e.g., see Figure 6, which incidentally also shows that our character art assets have improved marginally over time). This type of result indicates at least surface correspondence and helps to increase confidence in the workings of the PMF collection.

4.4 Integration Testing

In the efforts described thus far, a significant amount of time was invested in the development of a generic, reusable agent framework and to build up relatively valid synthetic agents (terrorists, opponents, security forces, crowds) to simulate a few example scenes (checkpoint crossings, protests, riots, looting, etc.). Substantial effort is necessary to cull various relevant sources and assure that value trees and other parameters lead to reasonably valid and correspondence-tested behavior. As these assets continue to develop and expand, certainly it would be advantageous to have the capacity to make use

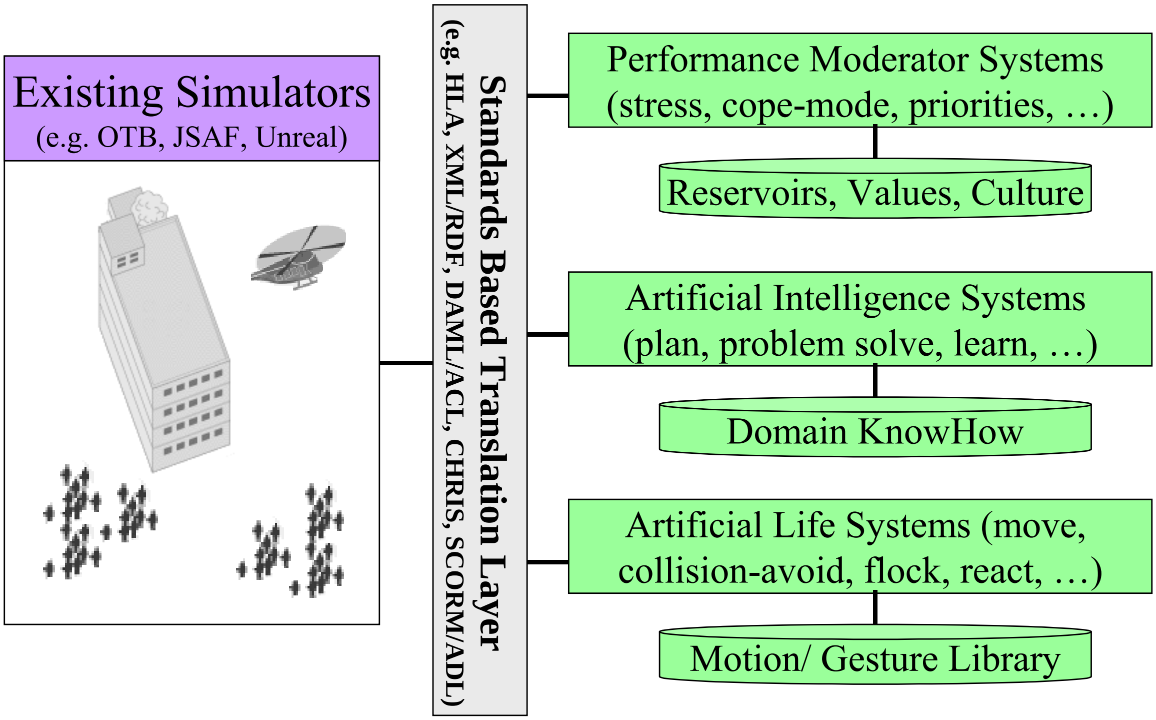


Figure 7. Human performance simulators present a need for interchange between diverse behavior modeling assets

of them in other simulators and to increase the realism of other characters in other synthetic worlds.

The PMFserv is not tied to a particular simulator. The examples presented here were run on simulators created by students, but the characters could be used in other simulators as well. It is intended that PMFserv should eventually become a resource from which simulation developers can 'drag-and-drop' agent minds onto other agent bodies in their own simulations or apply specific PMF components as needed to moderate the behavior of their own simulated cognitive sub-systems. In short, if a given simulator manages the bodies of its agents, a package such as PMFserv can provide the minds for those agents. (Simulator engines generally also animate terrain, buildings, vehicles, and physical space as is suggested on the left side of Figure 7.)

The right side of Figure 7 illustrates the claim made at the beginning of this chapter, i.e., that artificial life systems manage low-level functions and artificial intelligence manages rational reasoning functions. Thus, artificial life functionality is vital for improving the realism of kinesthesis, movement, and gestures of soldiers and non-combatants moving through space. Likewise, realism is improved by adding some form of artificial intelligence for higher-level mental functions such as, for example, vigilant decision making and tactical and strategic planning. Likewise, PMFs constrain the characters' intelligence and life functions as dictated by items such as fatigue, stress and coping levels, cultural standards and individual emotions.

Human performance simulators need to use all of these capabilities in combination. The center of Figure 7 suggests that one way to achieve this is to attempt to create a translation layer that is a set of interchange standards between the various modules. In the best of all worlds there would already exist human modeling interchange standards. At present, such standards are still in early development (e.g., HLA, DAML/OIL, W3C's human ML, XML/RDF, ADL's SCORM, etc). Behavioral interchange standards that would facilitate such interchange efforts do not yet exist; we are still in the process of deciding what such standards should be developed (Bjorkman, Barry, and Tyler, 2001).

The initial testbed for this effort is a multi-group project lead by the Institute for Creative Technology (ICT) of the University of Southern California, and also including Biographics Technology, Inc. (BTI), the University of Pennsylvania, and the Institute of Defense Analysis (IDA) (Toth, Graham, Van Lent, et al., 2003). The entire testbed is based on a "player" who uses the help of three Ranger-bots to secure a crashed helicopter in a Mogadishu-style crowd and militia situation. Michael Van Lent at ICT has developed a scenario and an architecture that uses the Unreal Tournament (Infiltration Module) game engine as simulator (www.epicgames.com) and that we all are 'plugging into'. Unreal is a 3-dimensional, first-person shooter style game engine. In the test scenario, crowd and militia bots are primarily PMFserv-controlled. SOAR supports the decision making of the Ranger-bots, which eventually might also be moderated by PMFserv. AI-Implant is an artificial life package that is used to manage art resources and provide low-level implementations of actions (e.g., navigation, movement). Finally, Unreal itself includes artificial life functionality that can be invoked and contrasted to those of AI-Implant.

By exploring ways of tying these systems together, we expect to increase our understanding of the requirements for integration. For example, we hope to answer the following questions, among others: How should the diverse human modeling systems interact (right side of Figure 7)? Can we get away with a socket-based message passing system, or will the real-time nature of the system require us to use API calls for speed (center of Figure 7)? How many agents can be supported at once without degrading simulator performance (left side of Figure 7)? It is hoped that this demonstration will set the stage for future integration efforts with real-world simulators and provide valuable insight into the requirements that must be met for behavioral interchange standards.

5. Conclusions and Next Steps

It is an exciting time in the field of human performance simulation due to the proliferation of methods that improve our capabilities and potential. Most simulation developers and sponsors are now working to extend their systems to permit interchange with other approaches and other vendors. As more of these types of interchanges are attempted, more will be learned. The enterprise of human performance simulation is too vast an undertaking for any one provider to have it all.

The purpose of this chapter has been to illustrate this panorama by exploring what is newly possible and identifying challenges that remain. Several lessons learned are worthy of brief summary review, as follows.

- The literature is helpful for improving the realism of behavior models. An in-depth survey of the literature shows that there do exist models useful to the development of cognitive models for synthetic agents. The problem we face is not a shortage of useful models, but rather the fact that such models have not yet been integrated. This chapter summarizes recent efforts to document available models, to determine how they might be integrated into a common framework, and to implement and assess the value of such a framework.

- Integrated models will improve the realism of simulated agent behavior. Efforts to model stress, emotion, and decision processes as integrated factors – as they are in real human beings – will present new possibilities for improving and expanding realistic synthetic agent behavior based on the interplay of multiple factors and settings. Training simulations will also benefit through the presentation of more realistic scenarios.

- Value sets are vital but require significant engineering. The approach presented in this chapter relies on a common mathematical framework (expected utility) to integrate many disparate models and theories such that agents can assess their value sets for goals, standards, and preferences and determine next actions they find desirable subject to stress induced limitations and bias tendencies. To apply this approach properly for any given simulation will also require extensive engineering to flesh out the lower levels of the concern trees. Our current efforts are aimed at adding a set of tools for authoring, maintaining, and visualizing psycho-social-physiological dimensions and assembling a reusable "cast" of characters to help speed future scenario construction.

- Emotion models are useful for culture-based utility and decision making. A related benefit of the approach presented here is its use of values-derived emotion to help generate utilities dynamically. In standard decision theoretic models there is no basis for agents to compute their own utility functions. Instead, these are derived by subject matter experts and inserted directly into the agent's decision module. By contrast, the approach postulated here requires subject matter experts to interact at an earlier stage of development, when they are needed to define underlying value sets from which synthetic agents derive utility functions, priorities, and tradeoffs. This approach frees experts from having to infer utilities, and it places the debate more squarely on open literature accounts of value sets and concern ontology.

- Interoperable human performance simulators are desirable and feasible. Useful complementary contributions have been made in the fields of artificial life, artificial intelligence, and performance moderators. Distributed computing technology today permits the interoperation and real time interchange of these complementary parts. One branch of the social agent simulation field has proclaimed the need to try and simulate at the simplest level possible (e.g., cellular automata agents that are at most 40 bytes of data each). Unless one must model large populations, there is little need to adhere to starvation diets such as this. Not much realism can be guaranteed through such an approach. The alternative presented in this chapter is, in effect, a "multi-agent agent." This appears to offer the most promising path toward creating agents that are realistic and valuable to trainees and analysts. The next order of business will be to scale the effort up to represent increasingly large collections of agents.

These conclusions portray a relatively optimistic picture. However, there remain several grand challenges. One of these is that although the existing behavioral scientific literature is vast, it is ill-prepared for and cannot yet be directly encoded into models that are useful for agent architectures. What's worse, most behavioral researchers focus on narrow PMF topics, largely neglecting the developer's need for integration. If the field of agent modeling and simulation is to reach its potential, it will need behavioral scientists to work toward the assembly of a fuller representation of factors that influence human performance.

Another grand challenge is the need for highly composable systems that allow scenarios to be generated on demand and just-in-time for the purpose of training and analysis. This is the "Holodeck" dream, which begs a flotilla of research and development priorities, only some of which have been addressed in this chapter. Related objectives might include:

- Shift attention from the development of automatons to the development of realistic agent behavior. Automatons ("bots") ignore constraints of physiology, motivation, culture, relationships, and standards-based conflicts that arise in the real world. When agents and situations are realistic (i.e., when they pass correspondence tests), this preserves immersion, and **greatly** increases training value and skill transfer.

- Assemble a reusable, easily-adapted library of realistic digital casts and avatars to populate a wide array of scenarios encountered by soldiers and police. These scenarios would include situations that involve civilian and collateral damage, battlefield clutter, asymmetric cells operating under urban settings, Operations Other Than War (OOTW), homeland defense, and a variety of other concerns and challenges faced in modern security and peacekeeping endeavors.

- Reduce, by at least an order of magnitude, the effort needed to introduce human performance modeling components (PMFs, AI, A-life, etc.) into simulators. Having a published interchange standard can be shown mathematically to guarantee this result: $O(N^2) -> O(N)$.

When and if we conquer these challenges, then it seems that several benefits will result for the state of the practice of human performance simulation. First, a sea change will arise in the field of psychological modeling, which will shift from a few hegemonic systems like SOAR and ACT-R, to a proliferation of collaborating best-of-breed PMFs, AI systems, and A-life components created by and widely shared amongst distributed researchers. Second, there will be few technological barriers to entry for crafting purposive behaviors of avatars, allies, crowds, opponents, digital cast extras, etc. A wide array of agent types with truly interesting and demographically- and culturally-validated behaviors will be added directly by "turn the dials" designers into videogames, movies, and analytical simulations. Third and last, this will lead to a leap-ahead capability for the field of complex systems analysis – rather than being reduced to studying swarms and cellular automata restricted to trivial rulesets, one could study emergent and evolutionary behaviors of large collectives in a deep way (nuances of personality, subtleties of culture, variability in desires, etc.). When the state of the practice shifts along these lines, we will then be comfortable saying that human performance simulation is a relatively mature field.

Acknowledgement

The PMF related research summarized here and PMFserv were supported by research grants from the Defense Modeling and Simulation Office (DMSO) and the Office of Naval Research (ONR), while the GSP Tree (emotion module) subsystem was supported by grants from the Ackoff Center and gifts from the General Motors Corporation. This research has benefited from the help of several teams of students in my courses and research staff – too many to mention by name. Further, I'd like to thank Joe Toth and John Tyler for many useful discussions about how to apply PMFserv. Any claims are the responsibility of the author alone.

References

Anderson, J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Angus, R.G., & Heslegrave, R.J. (1985). Effects of sleep loss on sustained cognitive performance during a command and control simulation. *Behaviors Research Methods, Instruments, & Computers*, *17*, 55-67.

Badler, N.I., Chi, D.M., & Chopra, S. (1999). Virtual human animation based on movement observation and cognitive behavior models. *Proceedings of the Computer Animation Conference, IEEE Computer Society, Geneva, Switzerland*, 128-137.

Badler, N., Palmer, M., & Bindiganavale, R. (1999). Animation control for real-time virtual humans. *Communications of the ACM, 42***,** 64-73.

Bjorkman, E.A., Barry, P.S., & Tyler, J.G. (2001). Results of the Common Human Behavior Representation and Interchange System (CHRIS) Workshop. (Paper 01F-SIW-117). *Proceedings of the Fall Simulation Interoperability Workshop, Orlando, Florida.*

Boff, K.R., & Lincoln, J.E. (1988). *Engineering data compendium: Human perception and performance* (HED Tech Report - 3 Volumes). Wright Patterson AFB, OH.

Bradshaw, J.M., Greaves, M., Holmback, H., Karygiannis, T., Jansen, W., Silverman, B.G., et al. (1999). Agents for the masses? *IEEE Intelligent Systems, 14* (2), 53-63.

Cornwell, J.B., Silverman, B.G., O'Brien, K., & Johns, M. (2002). A demonstration of the PMF-extraction approach: Modeling the effects of sound on crowd behavior. *Proceedings of the Eleventh Conference on Computer Generated Forces and Behavioral Representation, Orlando, Florida,* 107-113.

Crumley, L.M., & Sherman, M.B. (1990). *Review of command and control models and theory* (Report No. AD-A230 105). Washington: Defense Technical Information Center (DTIC).

Damasio, A.R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Avon.

Downes-Martin, S. (1995). *A survey of human behavior representation activities for distributed interactive simulation* (DMSO Client Report). Alexandria, VA: Defense Modeling and Simulation Office.

Driskell, J.E., Hughes, S.C., Guy, W., Willis, R.C., Cannon-Bowers, J., & Salas, E. (1991). *Stress, stressor, and decision-making*. (Technical report for the Naval Training Systems Center). Orlando, FL: Naval Training Systems Center.

EBAA'99 Program, http://www.ai.mit.edu/people/jvelas/ebaa99/ebaaprogram.html

El-Nasr, M.S., Ioerger, T.R., & Yen, J. (1999). A web of emotions. In J.D. Velasquez (Ed.), *Workshop Notes: Emotion-based agent architectures (EBAA'99), May 1, 1999, 3rd International Conference on Autonomous Agents (Agents'99).* Retrieved May 1, 2003, from EBAA'99 Program, http://www.ai.mit.edu/people/jvelas/ebaa99/ebaaprogram.html

Fudenberg, D., & Tirole, J. (2000). *Game theory* (7th ed.). Cambridge: MIT Press.

Funge, J. D. (1998). Making them behave: Cognitive models for computer animation. *Dissertation Abstracts International, 59* (12), 6378B. (UMI No. NQ33902)

Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gillis, P.D., & Hursh, S.R. (1999). Using behavior moderators to influence CGF command entity effectiveness and performance. *Proceedings of the Eighth Conference on Computer Generated Forces and Behavioral Representation, Orlando, Florida*, 237-251.

Gorman, P.F. (2000). The defence of Fombler's ford. *Army, 50*, 27-42.

Hammond, K. (2000). *Judgments under stress*. New York: Oxford Press.

Hayes-Roth, B. (1998). Animate characters. *Autonomous Agents and Multi-Agent Systems, 1* (2), 195-230.

Hendy, K.C., & Farrell, P.S.E. (1997). *Implementing a model of human information processing in a task network simulation environment* (Report No. DCIEM-97-R-71). Downsview, Ontario: Defence Research & Development Canada.

Horowitz, D.L. (2001). *The deadly ethnic riot.* Berkeley: University of California Press.

Hursh, S.R., & McNally, R.E. (1993). Modeling human performance to predict unit effectiveness. In D. Kamely, K. A. Bannister, & R. M. Sasmor (Eds.), *Army science: The new frontiers: Military and civilian applications* (pp. 309-328). Woodlands, TX: Borg Biomedical Services.

Janis, I.L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment.* New York: The Free Press.

Johns, M., & Silverman, B.G. (2001). How emotion and personality effect the utility of alernative decisions: A terrorist target selection case study. *Proceedings of the Tenth Conference on Computer Generated Forces and Behavioral Representation, Norfolk, Virginia*, 55-64.

Klein, G.A., Orasanu, J., Calderwood, R., & Zsambok, C.E. (1993). Decision making in action: Models and methods. Norwood, NJ: Ablex.

Labrou, Y., Finin, T., & Peng, Y. (1999). Agent communication languages: The current landscape. *IEEE Intelligent Systems and Their Applications*, *14* (2), 45-52.

Laird, J.E., Johnson, W.L., Jones, R.M., Koss, F., Lehman, J., Nielsen, P.E., et al. (1995). Simulated Intelligent Forces for Air: The SOAR/IFOR Project 1995. *Proceedings of the Fifth Conference on Computer Generated Forces and Behavioral Representation, Orlando, Florida,* 27-36.

Lazarus, R. (1991). *Emotion and adaptation*. Oxford: Oxford University Press.

McPhail, C., & Wohlstein, R.T. (1983). Individual and collective behaviors within gatherings, demonstrations, and riots. *Annual Review of Sociology, 9*, 579-600.

Nass, C. (1996). *The media equation: How people treat computers, televisions, and new media like real people and places.* New York: Cambridge University Press.

Ortony, A., Clore, G.L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Pew, R.W., & Mavor, A.S. (1998). Modeling human and organizational behavior: Application to military simulation. Washington: National Academy Press.

Rickel, J., & Johnson, W.L. (1998). STEVE: A pedagogical agent for virtual reality (video). In K.P. Sycara and M. Woolridge (Eds.), *Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98)* (pp. 332-333). New York: ACM Press.

Silverman, B.G. (1991). Expert critics: Operationalizing the judgment/decision making literature as a theory of "bugs" and repair strategies. *Knowledge Acquisition, 3*, 175-214.

Silverman, B.G., Johns, M., O'Brien, K., Weaver, R., & Cornwell, J. (2002). Constructing virtual asymmetric opponents from data and models in the literature: Case of crowd rioting. *Proceedings of the 11th Conference on Computer Generated Forces and Behavioral Representation, Orlando, Florida,* 97-106.

Silverman, B.G., Johns, M., Weaver, R., O'Brien, K., & Silverman, R. (2002). Human behavior models for game-theoretic agents. *Cognitive Science Quarterly*, *2* (3/4), 273-301.

Silverman, B.G., Might, R., Dubois, R., Shin, H., Johns, M., & Weaver, R. (2001). Toward a human behavior models anthology for synthetic agent development. *Proceedings of the Tenth Conference on Computer Generated Forces and Behavioral Representation, Norfolk, Virginia*, 277-285.

Sloman, A., & Logan, B. (1999). Building cognitively rich agents using the SIM_AGENT toolkit. *Communications of the ACM, 42* (3)**,** 71-77.

Tambe, M., Johnson, W.L., Jones, R.M., Koss, F., Laird, J.E., Rosenbloom, P.S., et al. (1995). Intelligent agents for interactive simulation environments. *AI Magazine*, *16* (1), 15-39.

Teigen, K.H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, *4* (4), 525-547.

Terzopoulos, D. (1999). Artificial life for computer graphics. *Communications of the ACM*, *42* (8)**,** 33-42.

Toth, J., Graham N., Van Lent, M., Alinden R., Silverman, B.G., Cornwall, J., et al. (2003, May). *Leveraging gaming in DOD modeling and simulation: Integrating performance and behavior moderator functions into a general cognitive architecture of playing and non-playing characters.* Paper presented at the Twelfth Conference on Behavior Representation in Modeling and Simulation (BRIMS, formerly CGF), SISO, Scottsdale, Arizona.

Virtual Naval Hospital website on battle stress at www.vnh.org.

von-Neuman, M., & Morganstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wayner, P. (1995). *Agents unleashed*. New York: Academic Press.

Weaver, R., Silverman, B.G., Shin, H., & Dubois, R. (2001). Modeling and Simulating Terrorist Decision-making: A "performance moderator function" approach to generating virtual opponents. *Proceedings of the Tenth Conference on Computer Generated Forces and Behavioral Representation, Norfolk, Virginia,* 39-44.