Departmental Papers (ESE)   Department of Electrical & Systems Engineering

4-1-2006

Human Behavior Models for Agents in Simulators and Games: Part II Gamebot Engineering with PMFserv

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

Gnana K. Bharathy
University of Pennsylvania

Kevin O'Brien
University of Pennsylvania

Jason Cornwell
University of Pennsylvania

Human Behavior Models for Agents in Simulators and Games:

Part II – Gamebot Engineering with PMFserv

Barry G. Silverman, Ph.D., Gnana Bharathy, Kevin O'Brien, Jason Cornwell
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

Many producers and consumers of legacy training simulator and game environments are beginning to envision a new era where psych-socio-physiologic models could be inter-operated to enhance their environments' simulation of human agents. This article explores whether we could embed our behavior modeling framework (described in Part I) behind a legacy first person shooter 3-D game environment to recreate portions of the Black Hawk Down scenario. Section One amplifies on the inter-operability needs and challenges confronting the field, presents the questions that are examined, and describes the test scenario. Sections 2 and 3 review the software and knowledge engineering methodology, respectively, needed to create the system and populate it with bots. Results (Section 4) and discussion (Section 5) reveal that we were able to generate plausible and adaptive recreations of Somalian crowds, militia, women acting as shields, suicide bombers, and more. Also, there are specific lessons learned about ways to advance the field so that such inter-operabilities will become more affordable and widespread.

Keywords: human behavior models; culture and emotions; simulator and agent interoperability; composability

1) Introduction

Today's world is on the verge of an era of ubiquitous agents – autonomous characters that assist in all endeavors at work, at home, online, in games, and in social settings. Yet today's agents are too easily perceived as mechanistic automatons, causing users to experience frustration, inappropriate expectations, and/or failures of engagement and training. Reliable pathways for creating more realistic and believable agents could ultimately help reduce barriers to interacting with as well as to creating behaviors of empathetic avatars, electronic training world opponents and allies, digital cast extras, wizard helper agents, and so on.

This is no where more apparent than in the military modeling and simulation community which is demanding human behavior models (HBMs) to satisfy a wide and expanding range of scenario concerns. Their interest goes beyond mission-oriented military behaviors, to also include simulations of the effects that an array of alternative diplomatic, intelligence, military, and economic (DIME) actions might have upon the political, military, economic, social, informational (psyops), and infrastructure (PMESII)

dimensions of a foreign region. The goal is to defeat adaptive foes adept at using local PMESII effects to their own advantage: e.g., see Runals (2004).

If the military is to have realistic and reliable models of the effects of DIME type operations upon PMESII dimensions, one must find ways to integrate scientific know-how across many disciplines. As the top of Figure 1 shows, science tends to be reductive, specialized, and siloed. Labs that study sleep deprivation don't also study impacts of non-lethal crowd control methods, and those specialists know little about political coalition dynamics. Yet, each of these, and more disciplines have something of value to contribute if we are to realistically model the type of effects just described.

Part I of this article presented a unified architecture for human behavior modeling that seeks to straddle and synthesize models and principles from physiology/stress, personality/culture/emotion, social/political, and cognition and perception. This is an approach to help modelers cull scientific models and first principles from the behavioral literatures so they can be edited, tested for their validity, and used to improve realism of agent behavior. Obviously, many efforts such as this effort are needed to make progress. Science continually must go through periods of synthesis across disciplines in order to uncover its shortcomings and to regenerate. This is the feedback loop that the right side of Figure 1 shows from synthesis to further empiric and reductive investigations. The current push for better models is uncovering and fueling many such studies at present. It is thus a productive time to examine synthesis of HBMs and methods for doing so.

Our computer implementation of the unified behavior architecture, PMFserv, provides one starting synthesis of models and principles. The current article, Part II, serves as an existence proof that this implementation can be harnessed and used to enhance agent realism and to help model and simulate certain pre-, during-, and post-conflict situations in other cultures. Since this is a case study, the answers we uncover will be largely limited to one instance, and not generalizable without further investigation. Also, no one HBM is sufficient to address all the concerns, so the bottom of Figure 1 also lays out a methodology in four boxes that raises the idea of federating other models as well. This vision leads to three sets of questions we explore in this paper:

1) Are models drawn from the literature useful and usable as agent minds? To what degree will they elevate an automaton into a realistic agent? Under what conditions do these models help agents pass (fail) correspondence tests?
2) Is the legacy simulator community (military and entertainment) ready and able to accept such plug-in models for updating the minds of bots that already exist in their software? If not, what obstacles exist and what fixes appear warranted?
3) What is needed to improve the composability situation so that digital casts can be created? From a knowledge engineering perspective, how do various methods and approaches impact affordability?

The motivation behind these questions is to explore if it is reasonable to federate models to foster composability. There is study after study that shows the lack of credible behavioral capability of the legacy systems (e.g., see Pew & Mavor, 1998; Anon., 1995; Bjorkman & Blemberg, 2001, among others). A federation approach could help to preserve the investment in legacy simulator and game environments, while making newer

character simulations and behavioral model innovations available. This path has been advocated by the Department of Defense, among others, who has identified a need for interoperability of human behavior models to help improve the realism of agents in legacy simulators: (e.g., see Finerman et al., 2001; Bjorkman, Barry, & Tyler, 2001; Toth et al., 2003).
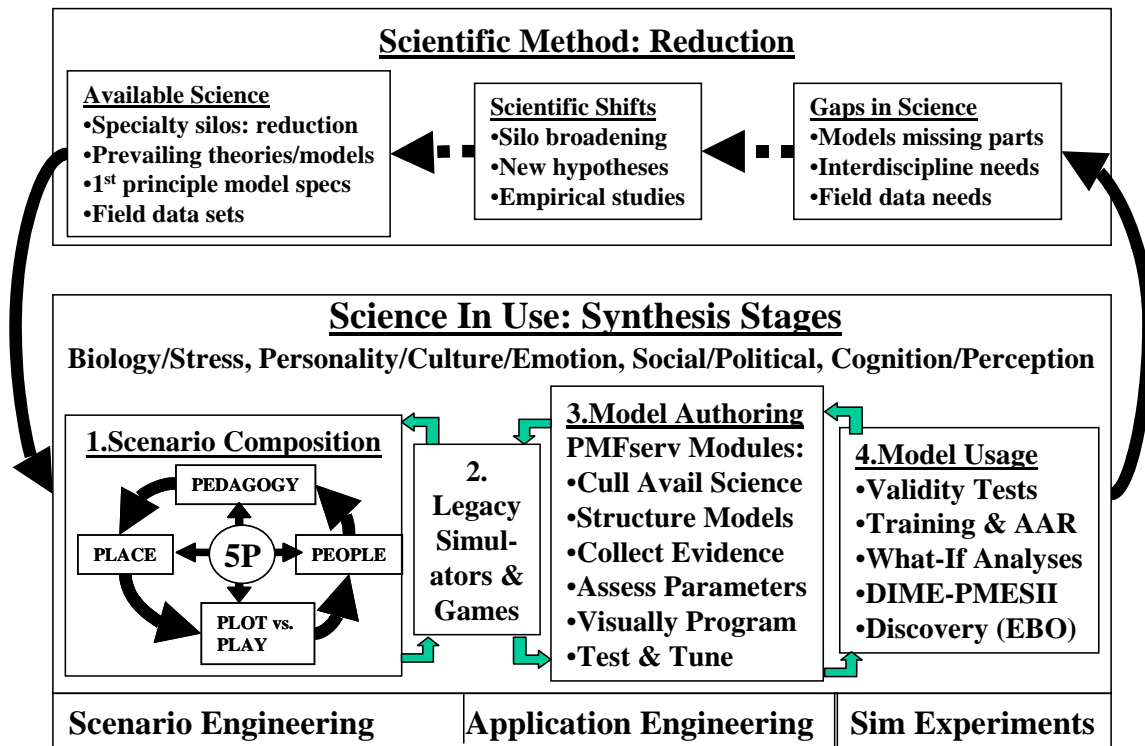
Scientific Method: Reduction

Available Science
- Specialty silos: reduction
- Prevailing theories/models
- 1st principle model specs
- Field data sets

Scientific Shifts
- Silo broadening
- New hypotheses
- Empirical studies

Gaps in Science
- Models missing parts
- Interdiscipline needs
- Field data needs

Science In Use: Synthesis Stages

Biology/Stress, Personality/Culture/Emotion, Social/Political, Cognition/Perception

1. Scenario Composition

PEDAGOGY
PLACE 5P PEOPLE
PLOT vs. PLAY

2. Legacy Simul-ators & Games

3. Model Authoring
PMFserv Modules:
- Cull Avail Science
- Structure Models
- Collect Evidence
- Assess Parameters
- Visually Program
- Test & Tune

4. Model Usage
- Validity Tests
- Training & AAR
- What-If Analyses
- DIME-PMESII
- Discovery (EBO)

Scenario Engineering | **Application Engineering** | **Sim Experiments**

Figure 1. The Four Stage Synthesis Methodology and How It Integrates New Science and Legacy Software into Human Behavior Modeling

The four numbered blocks in the Synthesis portion of Figure 1 represent a four stage methodology that we have evolved through several studies and that is the organizing framework of this paper. Frequently, the client has only a top level notion of the scenario to be engineered. For example in this case study, in the summer of 2002, the DOD/Defense Modeling & Simulation Office (DMSO) wanted to see if our PMFserv agent behavior framework could successfully run the local crowds and militia of a recreation of the Black Hawk Down scenario. To help the client develop their scenario further, we use a process labeled 5P (1st stage in Figure 1) and explained more fully in Section 1.1.

As part of the case study, the client also requested that we attempt to embed the PMFserv agent minds behind a pre-existing simulator. This is question set 2 above, and it is the nature of HBM today that one often must embed behind a client's legacy simulator. This 2nd stage of the methodology is a challenge. In a recent survey of five legacy combat simulators (JSAF, ModSAF, OneSAF, DISAF, JCATS), it was found that (1) one often can't discover if a given behavior exists or what level of fidelity its modeled at; (2) the software is growing constantly; (3) verification and validation needs of the legacy

3

software make it prohibitive for anyone other than the prime contractor to add updates (LaVine et al., 2002). This study indicated the need to find novel ways to off-load behavior modules and agent software to external servers where they can be separately maintained and validated. When needed they could be dynamically federated (i.e., interoperated) through a mediating service. This case study is one such federation.

As a result of these types of constraints, there is often a give and take negotiation where the scenario is altered to suit the legacy codes and/or the choice of legacy system is altered to support more of the scenario questions of interest. This negotiation also involves Stage 3 and the tradeoffs of what behaviors to model as well. For example, in our case study, we spent several months with our client and an integrating contractor investigating numerous legacy simulators before settling on the one described in Section 2 of this write-up. In the effort to clarify implementation details, Section 2 treats this decision as already completed, but it is an important stage of the methodology.

The third stage of our methodology of Figure 1 as already mentioned, consists of behavior Model Authoring. The six steps listed inside it are explained in detail in Section 3 of this paper. Sections 4 and 5 address the Model Usage stage of Figure 1. There we present results and findings of our Mogadishu correspondence test, though as Figure 1 suggests there are many other types of usage one could support beyond what was asked in this case study.

1.1) The Test Scenario and the 5 Ps

The sponsor of the test scenario (DMSO) with the help of our 5P approach (about to be defined) and their technical representative (IDA) posed a detailed Mogadishu recreation scenario for the purposes of testing the capabilities of PMFserv as well as for illustrating its potential for integration into other simulators. In general, scenarios are like stories and for that one invariably must define the components of and interactions between People, Place, and Plot. Since gameplay is involved, a 4th P (that of Play) is also included. Finally, since the goal is a training game, one must also factor in the pedagogical or training objectives (in analytical studies, these may be the policies that certain agents are expected to uncover). This section explains the Plot, Plan, and Pedagogical goals of the scenario. It also overviews People and Place, a topic we examine more in Section 3.

The scenario test was intended not just as a test of PMFserv, but also as a test of several other human behavior models (HBMs) as well.
- a traditional AI system for representing and reasoning about combat knowledge (Soarbots from University of Michigan). There are pre-existing Soarbots for Unreal that have significant rulesets for soldier operations and combat.
- a module for enhancing the physics and animation believability of the legacy world's embodied agents (AI Implant from BTI). 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). Unreal itself includes artificial life functionality that can be invoked and contrasted to those of AI-Implant.
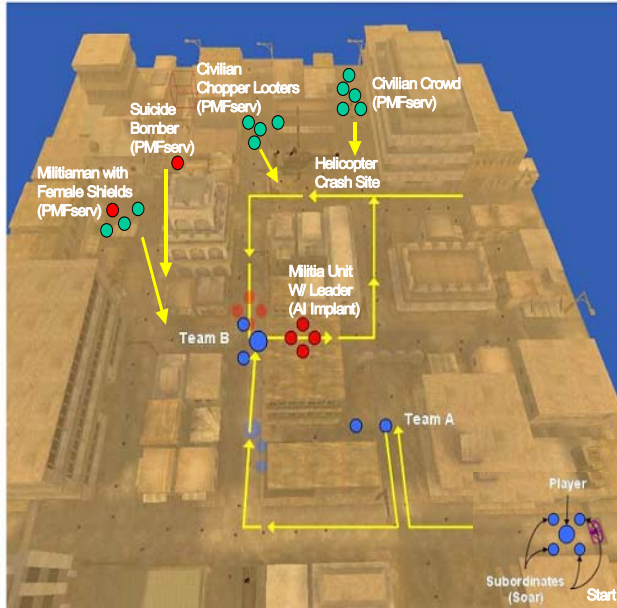- our PMFserv for managing the agent stress, emotions, and culture.

Figure 2. A Top-down View of the Terrain for the Mogadishu Test-Bed with the Starting Locations of the Soar (blue), AI.Implant (red), and PMFserv (Green) Agents

Various configurations were considered for the initial testbed, including the idea that all three agent modules might be integrated into the mind of each bot in the gameworld. In the end, it was decided that the first trial of this architecture should involve each agent in the gameworld being governed by a single HBM. Below we review how many agents are under the control of each HBM.

The test scenario and the testbed for this effort was 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 for Defense Analyses (IDA) (see Toth et al., 2003). The current paper primarily examines the issues of the PMFserv connection to the Interchange and to the legacy system. For an overview of the results across all groups, see van Lent et al. (2004a).

Custom art assets have been developed including terrain, buildings, and 3D models and textures for soldiers and weapons. The terrain consists of approximately 16 city blocks in a 4x4 street grid (see Figure 2). These blocks consist of interspersed multi-level buildings, obstacles, and a series of alleys.

In the Mogadishu scenario, a squad of four U.S. Army Rangers (one of whom is the player or trainee) deboard their Humvee on the bottom right of Fig. 2. Under the command of the human player, the squad then traverses the streets of Mogadishu in an attempt to locate a downed Black Hawk helicopter that they must clear of looters, destroy, and return safely from. Along the way, they encounter a variety of asymmetric threats and civilian crowds, each of which must be dealt with appropriately. More precisely, there are four AI.Implant militia that ICT implemented patrolling the middle of the level as militia. As the player emerges from the middle, the PMFserv controlled

"bots" begin to be encountered. From here onward, about 2 dozen PMFserv controlled bots populate the world as the Somali civilians (males and females) and Somali militia members. Also a terrorist bomber emerges.

In terms of pedagogical goals for the PMFserv gamebots, as the player and his subordinates advance upon the Durant Crash Site, they encounter two groups of PMFserv civilians, one gathered around the helicopter and the other looting inside it. The player and his Rangers (Soar) must encounter and disperse a crowd of Somali civilians both inside and outside the helicopter. In general, these Somalis have grown up with violence and should not be easily intimidated. Further, they must recognize when Rangers are vulnerable to swarming behaviors such as when a Ranger is alone, or his weapon is out of ammo.

If the player or Rangers kill a civilian, this should precipitate all males (and possibly a female) to feel so violated they will search for a way to revenge themselves on the Rangers. In many cases this should result in them appearing to flee, when in fact they are locating a weapon and intending to return fully armed and ready to engage. Also, the player and his Rangers must encounter a crowd of civilians with a Somali Militia shooting from behind them. The women bots have to make a decision to act as shields or not for the militia man. If they do act as shields, the militia's tactics should be to try and get the Ranger to kill one of the civilians. If the player or Rangers kill a civilian, this should precipitate a second threat which is a suicide bomber who appears as any other civilian male and is undetectable except that he advances without halting.

2.0) Testbed Architecture and Engineering

This section presents the architecture and software components needed in order to implement the PMFserv portions of the test scenario. There are many possible ways to create a federation of models. The center of Figure 3 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, & Tyler, 2001). However, in our effort we wanted to explore what such standards might need to include, and we will say more about this in the discussion.

As the left side of Figure 3 illustrates, the architecture includes the legacy game/simulator environment of the client. The middle of Figure 3 includes some "standards-based" form of interchange. Finally, the right side of Figure 3 shows the PMFserv and its related services act as the gamebot server. The bots on the client side implement and illustrate the agent bodies, actions, and results, while the server side provides the agents' motivations, stress, coping style, emotions, personality, and decisions. The next three subsections provide more detail on these three components, respectively.
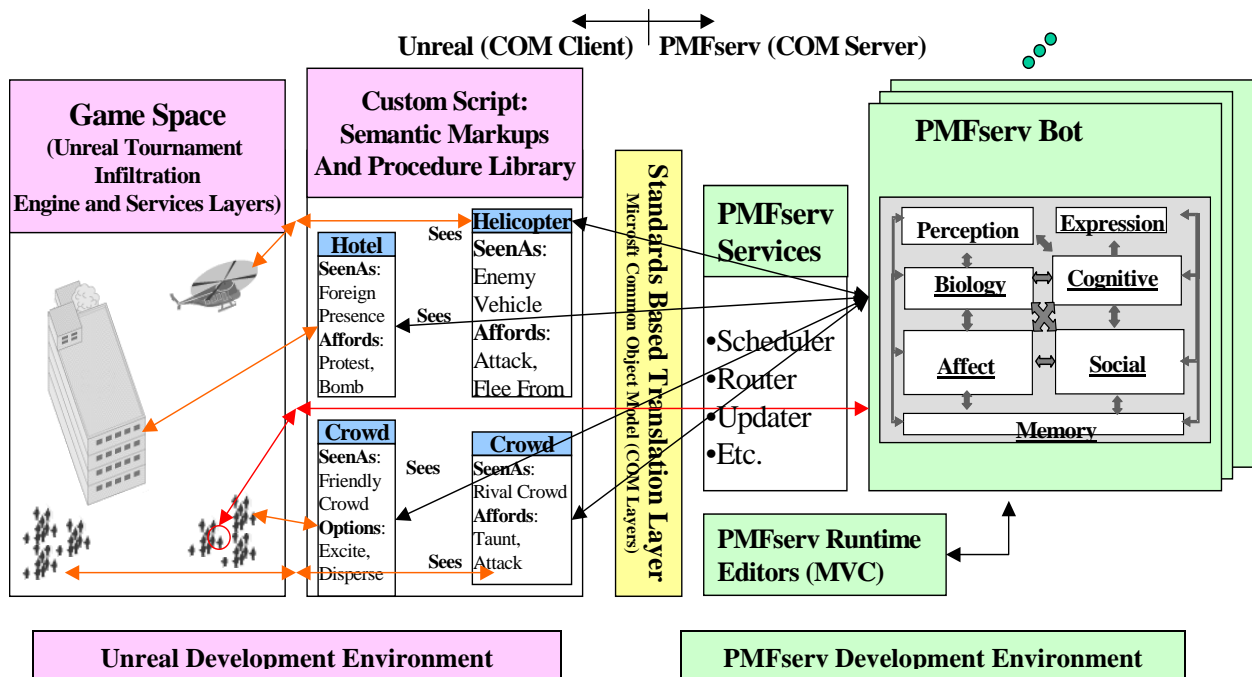
Figure 3. Architecture Adopted for Interchange and Integration of PMFserv with Unreal Tournament

2.1) Simulator on the Client Side: Unreal Tournament - Infiltration

Unreal Tournament (UT) is a popular First Person Shooter (FPS) game, released in 1999, that includes one of the most widely used interfaces to allow hobbyists and developers to extend and adapt (or "mod") the game to meet particular needs. The UT Game Engine (UTGE) is the driver behind any game or simulation scenario developed in UT. Through the mod interface, many of the UTGE components have been "exposed" giving hobbyists and developers a consistent programming interface to make changes to many aspects of the existing game (rendering, physics, AI, networking).

The off-the-shelf version of Unreal Tournament is itself not a realistic simulation of urban combat. However, a mod called Infiltration modifies UT to include more realistic soldier and weapon models (such as the M16, the M4, and the AK47), base-level behaviors, and tactics. The character models resemble soldiers and civilians. Infiltration provided the baseline character movement (walking, running) and weapon handling (firing, reloading, unjamming) actions. ICT enhanced the Infiltration mod with the custom urban terrain, but there were no custom character models representing Somali civilians. Those need to be created as delineated below.

PMFserv bots are mind, and not body. Thus they need skins, bodies, physics, kinematics, animations, etc. provided for them from the game engine's existing bots. If a PMFserv bot decides to observe, flee, taunt, loot, flock or swarm with the crowd, attack, die, etc. there must be game side code to execute and animate these actions. For a successful PMFserv demonstration the most important capability is the ability to represent changes in the mental and physiological states of our agents in the 3D models they are controlling.

Figure 4. View of Some of the PMFserv Controlled Bots in the Unreal-Mogadishu Environment

This translates into a variety of models, skins, and animation cycles for each agent type. The artist/animator was contracted to provide the skins, but this did not occur. Instead, we found two (break dancing) Somali-looking civilian skins in Unreal Tournament's public library that had far more simplistic behaviors than these, and with which we could only create a scaled down implementation of the crowd gestures and actions. These included a woman with a blue bourka and a male with red shawl and white robe (see Figure 4). These shareware bots existed with many of the low level behaviors including breathing, a celebratory animation that looks a bit like break dancing, running, picking up a weapon, shooting, dying, and the like. Many of these built-in behaviors had to be modified or overridden to slow them down and make them fit our needs. These bots did not include navigation routines, walking, flocking, swarming, attaching to crowds, taunting, and so on. They had no physiology in the sense of fatigue, noise reactions, and so on. They had no emotions, coping styles, stress reactions, or decision making functions. Much needed to be done to finalize the bots for the scenario vignettes and game called for here. These changes were coded in Unreal Script and are shown in that layer in Figure 3. From the Somali bots depicted in Figure 3, we managed to cobble together and alter the break dancing and other animations so in the end the visual behavior of the bots loosely approximates many of the desired animations. One can see videos of these at www.seas.upenn.edu/~barryg/HBMR

2.2) The Interchange Layer

In the ICT testbed, the interchange between PMFserv and Unreal Tournament that most satisfied our timetable and budget limits was the Microsoft COM interchange standard. Since PMFserv is in Python which sits atop the C language and since UT runs in Windows for the Testbed, it was relatively straightforward to adopt and implement the Component Object Model (COM) specification and software from Microsoft (Williams & Kindel, 1994). COM refers to both a specification and implementation developed by Microsoft Corporation which provides a framework for integrating components. COM defines an application programming interface (API) to allow for the creation of components for use in integrating custom applications or to allow diverse components to interact. However, COM is a low level service and in order to interact, components must

adhere to a binary structure specified by Microsoft. As long as components adhere to this binary structure, components written in different languages can interoperate.

To use COM for our interchange required us to adopt a client-server approach (illustrated by earlier Figure 3) which required us to do the following:

- Create a COM server for PMFserv on the Python side that exposes itself via COM to any application that is COM-aware. This made use of a pre-existing freeware DLL or Python module for mapping between Python and Microsoft's COM library.
- Create a Dynamic Linked Library (DLL) designed to work with Unreal Script that turned Unreal into a COM client. This DLL was written in C++ and was inserted into Unreal as "native code".

This enabled Unreal Script to make direct calls to PMFserv functions and to send updates for specific bots. At runtime, Unreal operates a process with the Unreal-COM client as a sub-process. The PMFserv runs as a process on the same machine (currently) while the PMFserv COM Server runs as third process under the control of the Windows COM facility. This COM server has two threads, one ongoing thread that monitors client requests while the other thread is spawned when client requests occur and lives until they are satisfied from the COM server side.

Unreal Tournament Environment	Custom Unreal Script	COM Standard Interface	PMFserv Environment & Services	Individual PMFserv Bots
•Art assets •Weapons •Animations •Physics •Sound Effects •Standard Behaviors •Basic Bot AI •Game Engine •Camera & Display Services	•Overrides of Unreal Behaviors •New AI & Behaviors •Semantic Markup of World Objects & Events	•COM client in UT (C++) •Microsoft's COM in Windows •COM server in Python	•Event Data Router •Semantic Labeling of Events •World Object Affordances •Bot Responses Uploader	•Event Sensor •Memory Unit •Physiology/ Stress Module •Personality, Culture, & Emotion Module •Decision Module •Response Selector

Table 1. Functionality Allocations across the Runtime Interchange Protocol Using COM to Connect PMFserv to Unreal Tournament

As Table 1 shows, there are essentially seven layers to this protocol – two for UT, three for COM, and two for PMFserv. Thus if an event happens in UT, it must be sent through all these layers for the relevant Bot in PMFserv to sense it and formulate a response. A similar path must be traveled in the reverse order for the response to reach UT and be played out by the UT game engine. The bots in PMFserv cannot directly call Unreal functions, but instead can poll the PMFserv Services Layer to find out if anything has been updated since the last tick. Currently PMFserv operates on the same machine as Unreal, however, the interchange makes it straightforward to provide parallel processors,

and by that to increase the number of bots in Unreal without adversely affecting performance.

This seven-layer protocol sounds potentially complex, yet it performed quite well in practice and did not lead to latency of note in the responses of the bots.

2.3) Server Side: PMFserv
PMFserv was described in detail in Part 1 of this article so there is no need to repeat that here. What is new here is the second to last column of Table 1, which is an expansion of the "services" block of Figure 3. Many of these services are simple synchronization, router, and uploader types of functions. One interesting service is the semantic mark up and affordance objects. These were introduced in Part I and we will discuss them further in Section 3.2. With the integration issues now out of the way, it is possible to focus on PMFserv and UTI as a single environment as the next section will proceed to do.

3) Agent Behavior Model Engineering

At this point we return to the issues of how to bring scientific principles and models, where available, to bear so as to enhance the reliability and realism of the agent behaviors. This corresponds to the third stage of our methodology from Figure 1, the block labeled behavior 'Model Authoring'. This stage consists of six steps we explore in this section. Before doing so, we should mention that following these steps does not preclude using other methodologies. Rather, we believe that many methods exist for amplifying the 5P approach and the six steps explored here. Thus we make use of other methods as needed, such as human behavior modeling (cognitive task analysis, protocol collection, personality instruments, etc.); social simulation design methodology (Gilbert, 1999); instructional design methodology (Gibbons et al., 1998); game design (Fullerton et al., 2004); knowledge engineering (Schreiber, 1999); and object oriented software analysis (Jacobsen, 1992), among others. However, none of these alone provides a clear path through the stages and the steps we enumerate in this article.

We go through the six steps of this behavior authoring stage for each module of PMFserv. Thus as a first pass on the Mogadishu case, the 5P process and some initial literature collection reveal that we need to model the following, subject to limits of the animation environment:
- **Archetypes** -- Four kinds of archetypes are needed including civilian looters/observers who can turn combatant, militia who can act as suicide bombers, females as shields, and some clan leader types.
- **Biology/Stress** – reservoirs and settings for exertion, wounds, adrenaline, effects of chewing the Khatt weed, multiple gunshot wounds required to kill them, round the clock effects/fatigue, event stress, time pressure, and emergence of coping modes such as unconflicted adherence, vigilance, and panic, among others.
- **Personality/Culture/Emotion (Values and GSP Trees)** – Goals, Standards, Preferences trees of members of the Habr Gidr subclan that capture values about belonging to family/clan, devotion to cause, jealousy of America, hatred of Rangers,

impact of seeing Rangers as vulnerable, impact of seeing loved-one wounded/killed, fighting ferocity/willingness to die, switching to combatant, treating women as objects, etc. (e.g., see Farah, 2000; Hussein, 1997; Abshir, 1998).

- **Social** – Clan alignments, inter-personal attachments, communicating, shared distrust, helping each other, effect of mob-rule (e.g., taunting, flocking, advancing, swarming, rioting), acting as human shields, converting identity to combatant, and dragging/stabbing the dead enemy (e.g., see Bowden, 1999).
- **Cognitive –** Civilians able to select a range of choices like observing, curiosity/attraction to noises/key sites, looting, fleeing, etc. The trained militia being smarter than the civilians-turned-combatants (CTCs) -- shoot more accurately, avoid killing each other in cross-fire, use women for cover, and so on (e.g., Lewis, 1994; Bowden, 1999).
- **Perception –** Physical situation, verbal communication objects, intent of others.

For each module of PMFserv, the six step authoring methodology helps to flesh out the specification, organize domain knowledge, and bridge the divide between specifications and software coding and tuning. Following this approach, all of the PMFs in the list above were implemented for the Mogadishu scenario. The next two subsections illustrate this for two sample modules. Due to space in this article, the interested reader is referred to Silverman et al., 2003 for a more complete treatment of the rest of the list. Also, Bharathy et al. (2003a, 2003b) present research done with a trauma surgeon to develop and tune the biological module (e.g., exertion, multiple types of wounds, stimulants, etc.).

3.1) Authoring the Agents' Personality/Culture/Value Trees

Our approach to modeling value systems was explained in Part I as driving an agent's cognitive appraisal, affect, and emotions. This is where we find much of personality and culture taking hold. Specifically, in PMFserv, this is modeled via three sets of trees called short term <u>G</u>oals, <u>S</u>tandards for behavior of self and others, and long term <u>P</u>references, or GSP trees. The steps needed in this module are to author each archetype's GSP trees.

Step 1 – cull the science: Consideration of cultural differences is not a new research topic, though there is little consensus on how to model culture. An early researcher, Hofstede (1980), contributed individual differences for five cultural factors. Though those were derived from international workers and include scores for groups around the world, they don't clarify how a factor will translate into an agent action, something we need in order to put this idea to use. Nisbett et al. (1999), in turn, focuses on cognitive processes and how Far Eastern vs. Western cultures change their perception and processing. While intriguing, it is not clear which of these two cultural poles, if either, applies to Somalia. Eidelson and Eidelson (2003) suggest that there are key beliefs or "dangerous ideas" that individuals and groups hold. These personal mindsets and collective worldviews can also differ culturally. Finally, Feltovich et al. (2004) define culture as systems of regulation external to the individual agent, including formal laws, religious tenets, and norms of practice. These create order within groups and define the standard options available to members. At the time of our test case, none of these cultural influences on adversary intent and behavior had yet been adequately represented in computational models, and the latter had not yet been published. If we were to do this study again, we would use more of the Eidelson and Feltovich suggested approaches, and

indeed in our newest work we are implementing some of these ideas collaboratively with those authors. However, our approach in 2003 may be equated to straddling that of the latter two, though rather informally.

Step 2 – structure the models: GSP trees hold the values that the agent's emotion module applies to evaluate the state of the world and the actions of others. They also are the motivators for the agent's own actions, and give rise to event stress (failed values) as well as subjective utilities for next action choices. One wants to design them to make sure that each possible action in the scenario will impact some branch of one or more of the GSP trees. However, GSP trees are forgiving structurally and one can build in redundancies or contradictions without much penalty. The actual structures we derived for the GSP trees of our archetypes went through a number of iterations. A slimmed down version for this article (except the leaders) is shown in Figure 5. At the time of this exercise, we had built GSP trees for a variety of crowd scenarios including domestic protests, the Intifadah, and soccer hooligans in the UK. Many of the archetypes in these crowd scenarios had similar structures in their GSP trees, although our leader structures tended to differ and still do. For example, we had a fair amount of success with Preference Tree structures indicating long term desires about locations, situations, and peoples, and that is reflected in Figure 5. Again, as Figure 5 also shows, we often used a Maslow type of structure for short term needs in the Goal tree, particularly for followers and cell members. Finally, we found that Standards trees tend to be ideally suited for adding nodes about types of actions that a group of agents is willing to take. Thus these structures tend to take on similarity to what both Eidelson and Feltovich refer to as internal mindsets and external norms of the group, respectively.
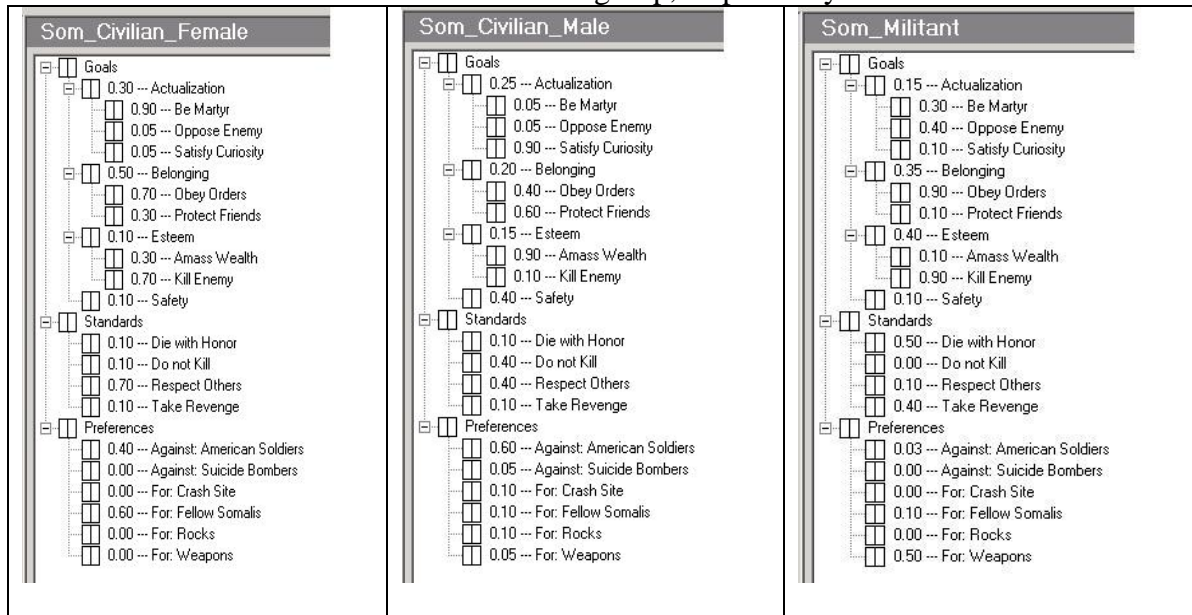

Figure 5. GSP Tree Structures for Three Archetype Agent Classes

Step 3 – collect evidence: Once the GSP tree structure is settled, one begins the process of assessing how important each branch is to a given archetype in the scenario. For the Mogadishu scenario, the data was available as empirical, narrative materials

consisting of a body or corpus of many statements of biographical information, and historic accounts (e.g., see Bowden, 1999, Farah, 2000; Hussein, 1997; Abshir, 1998). These empirical materials were organized into evidence tables through a modified content analysis process by breaking statements into simpler units with one theme (replicating statements when necessary), adding additional fields, namely reliability and relevance, and then sorting. For illustration, the following is an excerpt from the evidence table pertaining to the behavior of a Somalian woman.

Theme	Evidence	Reliability	Relevance
S1, G2	The females in Somalia suffer from inferior role in society and they often act in a subservient nature to men (Nelson, 1982)	Reliable & Generalizable	Training Data
S2, G3	Woman dart past driver's side of the truck and men told not to shoot because she had a kid. Woman turned and holding baby in one arm, took out a pistol with her free hand. (Anecdotal Pg 106, Bowden, 1999)	Reliable & Anecdotal	Training Data
S3, G3, G4	Woman began creeping up the alley directly toward the machine gun with guy hidden behind her. (Anecdotal Pg.43, Bowden, 1999)	Reliable & Anecdotal	Training Data
G2, S3	… her subordination to her husband is emphasized in the traditional beating that her husband is supposed to administer on the wedding night with a ceremonial whip. …… Throughout her married life a wife is expected to sustain this ideal of male domination, at least publicly…(Lewis, 1994, p.56).	Reliable & Generalizable	Training Data
S3	Somali man fired an RPG from behind a crowd of women. (Anecdotal, Pg.62, Bowden, 1999)		Testing Data

Table 2: Sample Evidence Table

In this fashion, a team of four undergraduates tackled the Mogadishu knowledge engineering as their senior design project (Lombardo et al., 2003). Each student researched the spreadsheets for the type of Somalian (female, civilian male, militia man, and clan leader) and markups of world objects from that Somalian's perspective. For each of their spreadsheet values, the students' spreadsheets include traces to actual interviews or literature sources that they felt justified their interpretation of that node or parameter setting. Over the past few years there have been a dozen student projects that successfully used this spreadsheet approach to produce term papers that cull references from the literature that support the various tree branches and weights assigned to bots of a given archetype and affordance levels for various world objects. Since spreadsheets are easily updated, they are a convenient knowledge engineering tool during the early stages of research and revision. These spreadsheets help to bound the effort and provide the knowledge engineer with a launching point for the subsequent steps of the process.

Step 4 – Assess Parameters: The next step is to assess the importance weights on each branch of the GSP trees. When the number of nodes to be compared increases, then assessment of weights is difficult without an appropriate technique. Such a weight assessment process is subjective, however, it is improved by pair-wise comparison using the Analytical Hierarchy Process (AHP) based scoring scheme (Saaty, 1982). Incorporation of an AHP-like pairwise comparison caters to the fact that at a given time,

the human mind can comfortably and reliably compare only two attributes. This also helps eliminate inconsistent ranking within the same groups, provides more systematic processes for assessment of weights, and leaves an audit trail in the process. The pairwise comparison assessment also takes into account the knowledge from differential diagnosis, using the ordinal rankings to crosscheck against the weights estimated. Let us look at the weight estimation for the standards tree for a female Somalian archetype. This process makes use of a format such as illustrated for GSP tree nodes in Table 4. Following this type of process, all relevant pairs of sibling nodes at a given level of the tree are compared and the weights for the GSP trees are enumerated. For example, in Table 3, when "Respect Others" is compared against "Die With Honor," the former was found to be strongly more important, giving a score of 7. If the order of comparison were to be reversed, it would be the reciprocal. The geometric mean along each row, when normalized gives the weights. The last column in Table 4 shows the finalized weights for the Standards Tree of Figure 5. In the same manner, the weights for all the GSP tree nodes were assessed for each subtree of each archetype.

Table 3: Questionnaire for Pairwise Comparison

......

Which of the following would be more important to a typical Somalian woman:

| Die with Honor | | & | Respect Others | | Not sure | |

....

How much more would she prefer that? [Or How much more important would this be to her?]

Equally	Slightly	Strongly	Very Strongly	Extremely
1	3	5	7	9

Table 4: Weight Estimation

	Die with Honor	Do not Kill	Respect Others	Take Revenge	Geometric Average	Weight
Die with Honor	1	1	1/7	1	0.61	0.10
Do not Kill	1	1	1/7	1	0.61	0.10
Respect Others	7	7	1	7	4.30	0.70
Take Revenge	1	1	1/7	1	0.61	0.10

Step 5 – Visually Program: With all the needed inputs now nicely estimated and organized, one next uses the visual editors of the PMFserv development environment to author each of the bots of that level including filling in their GSP trees (structures, nodes, weights). A similar exercise is done for other PMFserv modules to fill in physiology reservoirs, stress thresholds, relationships, and so on. To help with all these steps PMFserv includes a number of editors including bot and object creation editor, physiology editor, emotion module editor, decision editor, affordance editor, action editor, and others. Some of these editors were explained in Part I. Some of these editors are illustrated along with the results in the next two sections. Earlier Figure 5 shows the actual visual representation and editing environment for GSP trees in PMFserv.

Step 6 – Test and Tune the Models: In any of the modules of a PMFserv agent, one does not expect the prior five steps to lead to perfectly tuned models. Few of the scientific theories and PMFs being synthesized are mature and the process of integrating them does alter their original derivation as well. The implication is that one should explore the behavior in the neighborhood of the existing weights, with sensitivity analysis. In the Mogadishu case study, however, time and budget allocations from the

client limited us to only a few, manual investigations of the sensitivity surrounding the most sensitive weights. In this regard, the GSP weights are subjective estimates and are, hence, associated with uncertainty. While some behaviors are obvious, explainable and routine (observing strangers arriving, socializing), others might be key, critical (being responsible for a key change in the course of decisions) behaviors that might often come as a surprise. Examples of such key, critical behaviors include Somali women acting as human shields and militia carrying out suicide attacks against the troops. Given the significance of these events and the uncertainty associated with the estimates, it is appropriate to investigate the sensitivity associated with the key and critical behaviors. For example, the female's decision to act as shield derives significant utilities from such nodes as actualization (attaining Martyrdom) belonging (obeying orders and protecting friends) while failing on safety. The action is helped by the low weights on safety and esteem nodes that were derived in the Goal Tree for a typical Somali female. Clearly there is a point beyond which altering these weights would shift the female's tendency to act as a human shield. However, reducing a few obedience-related weights by 10% and using the difference to double the safety weight was attempted (recall that all weights of a given parent must add to 1.0 under Bayesian mathematics). This had no impact -- the females are slightly more individualistic but still provide cover with their own bodies.

3.2) Knowledge Engineering of the Objects (Perceptibles Markup)

In Step 5 from the previous section, one also performs another parameter editing activity that was glossed over there but that we focus on now, though we omit the detail of the steps for structuring perceptions, collecting evidence, etc. Specifically, one also uses the PMFserv editors to markup the major objects of the world with affordances – these are the actions one can take on world objects and the valence and intensity impacts those actions afford to the relevant leaf nodes of a given bot's GSP trees. Since each world object might be perceived in different ways at different times (e.g., as a flying helicopter, as a crashed helicopter) and in different ways by different bots (e.g., crashed helicopter is lootable vs. to-be-protected), the markups take some effort. However, this effort facilitates the affordance approach wherein the objects of the world contain the perceptions that we might ascribe to them. As with our 6-step process, the precursor to this effort is to fill out spreadsheets on the markups for each object from each observer's perspective. Once that is done, the affordance markups may be entered into the PMFserv environment. To help with all that, PMFserv includes a number of editors including bot and object creation editor, affordance editor, action editor, and others. Figure 6 illustrates some of these markup editors. The window to the left shows one militia male and three Somali women. These are marked up objects as well as PMFserv agents. As a markup, the females might see the militiaman as a fellow Somali or as a male demanding cover (upper right window). The male, in turn, might view the female as a fellow Somali woman or as human cover. They view women more as property objects than as humans (in their relationship slider low "Agency" means objectifying a person). Also, the markup for the militia male, shown here in the lower right editor window of Figure 6 exposes the actions that one can perform as a female and what it affords to the viewer (the female). In this case, if she gives cover to the male and acts as a human shield, she receives a number of positive GSP tree activations as the small vertical bars on the right edge of each GSP

tree node indicates. Only the "Safety" and "Respect Others" bars are negative (red), while all others (protect, obey, die with honor) are positive. Also, "ticks" indicate how long the activations persist in the emotion model.
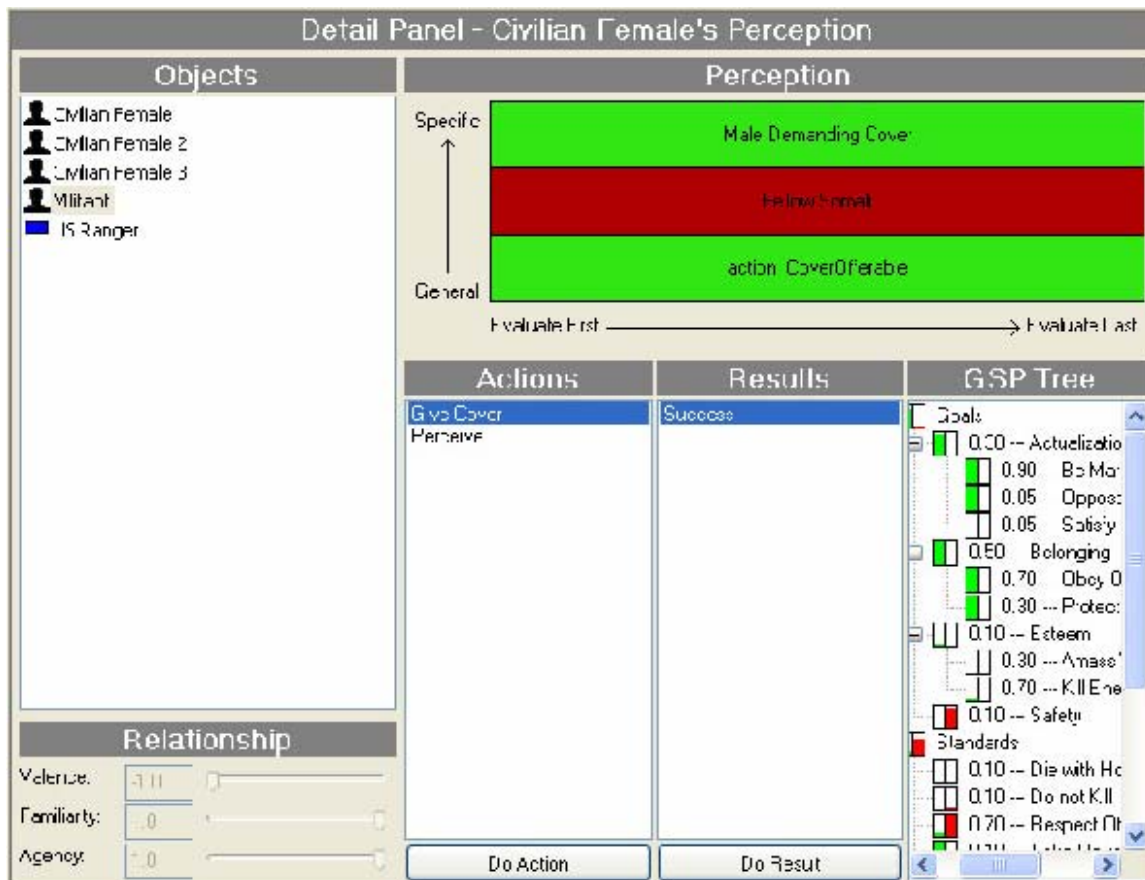

Figure 6. Editor Windows for World Markups and Affordances

4) Select Results from the Testbed

In this section we overview some illustrative results of the integrated Unreal-PMFserv testbed. Figure 7 displays the Physiology model where one can see that Stomach kcals contribute to the muscle Energy Tank which whose waste valve, in turn, is influenced by the Health and Sleep Tanks. Also, the Stimulant Tank can influence the Exertion Valve and how open or shut it is. In the first row, a Somalian male is shown who has some unfilled capacity in several of his reservoirs, but is unfatigued. We then force him to run around the virtual world until his stomach kcals and muscle or energy reserves are all drained. As the lower row shows, he is in a crouch, barely able to move and about to collapse from exhaustion. He is also gasping loudly for air in the demo. For tuning these parameters, one can set the controls for each of the reservoir thresholds as well as opening or flow rate of the drainage gates and valves as the right side of each display reveals. This can be done ahead of time or by pausing the simulation and in the middle of a test run. In this fashion, one can get the rates to a desired level. Tuning of the Mogadishu models was accomplished by Bharathy et al. (2003b).
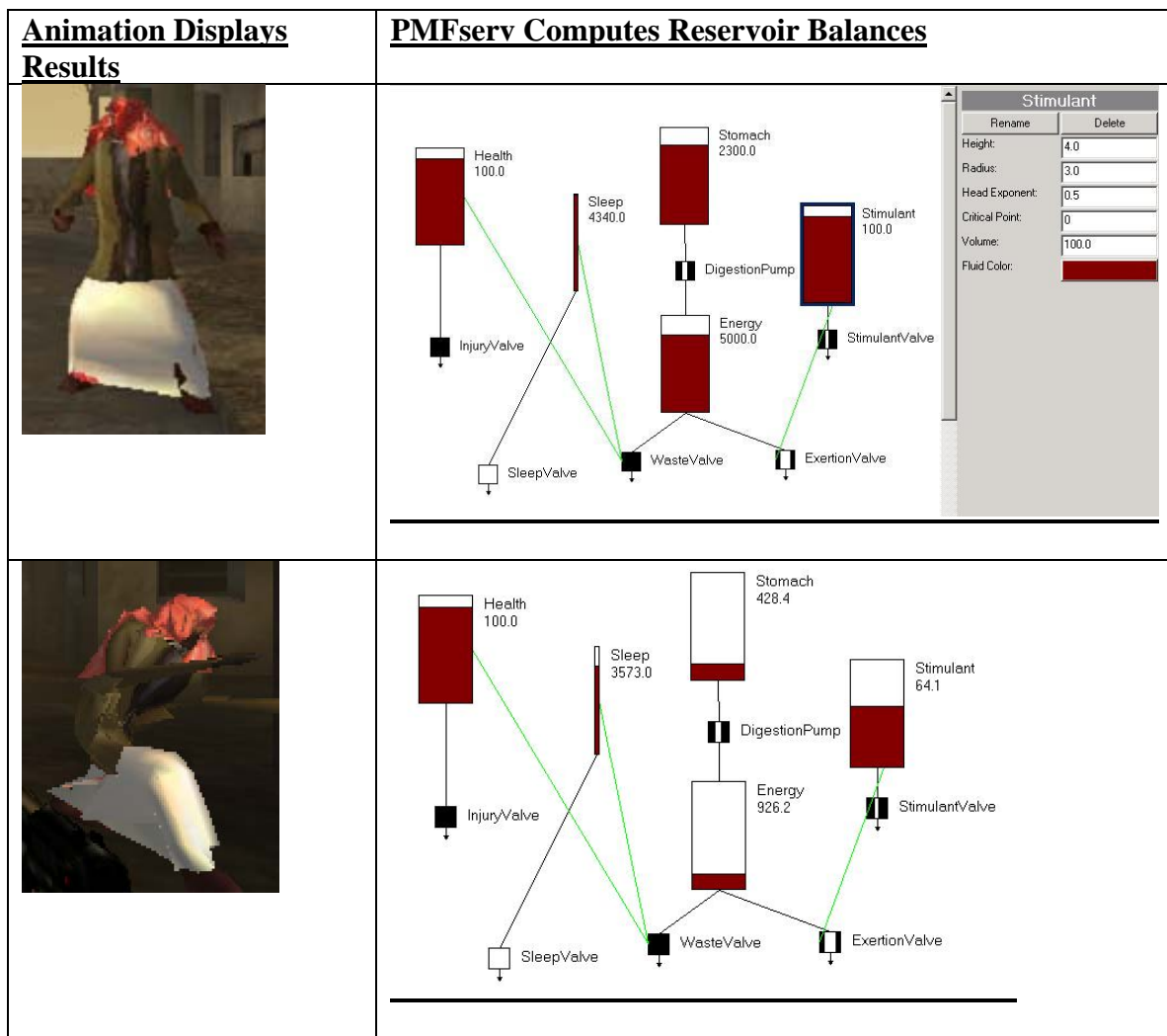
Animation Displays Results	PMFserv Computes Reservoir Balances

Figure 7. Somalian Male Before and After Physical Exertion

Figure 8 in turn shows a Somalian male being chased away from looting the helicopter. The player is the Ranger and his gun is in the foreground of the image, aimed at the looter and chasing him off. The display atop the image depicts the emotional state of the

17

looter at this moment (new emotions are generated with each new event in the simulated world). In the current time tick, all emotions are negative including distress over goals being thwarted, reproach against the Ranger, and anger for the situation. These emotions come about from the situation and due to his GSP tree weights. At the base of Figure 8 one can observe his Standards Tree and the weights. On the computer the branches are colorized as well – green for succeeding nodes, red when nodes are failing, and purple for mixed results (often in a parent with one success and one failed child). We can see that his standards for how people should behave include: do not kill, respect others, take revenge, and die with honor. Clearly, the Ranger has violated his first two standards by pointing the gun and chasing him off. Not shown are his Goal and Preference Trees which cause him to prefer free loot, and which in this case include a fairly high weight on the goal for safety. In the ensuing ticks this fellow moves back a distance to the safety of the crowd, and then goes home to retrieve his gun. While he is not in the militia, gun fights are a way of life in the Bakara Market (notice his low relative weight on 'do not kill'). As a result, some of the other looters are bolder than this one (lower weight on safety), and it takes more than just pointing your gun to chase them off.
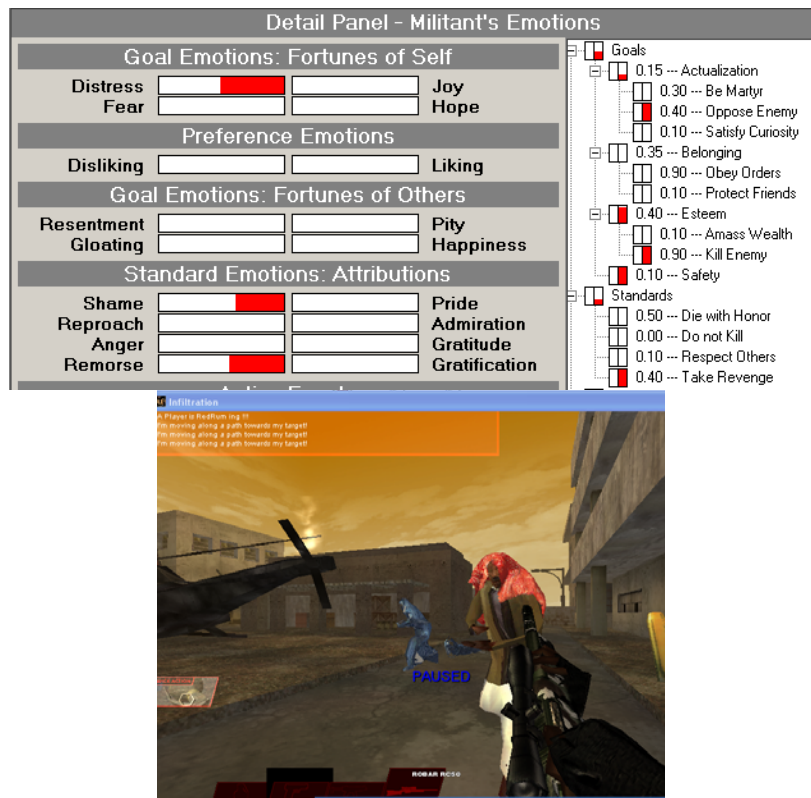

Figure 8. Emotion Display of Looter Being Chased Away from Helicopter by Ranger (Player)
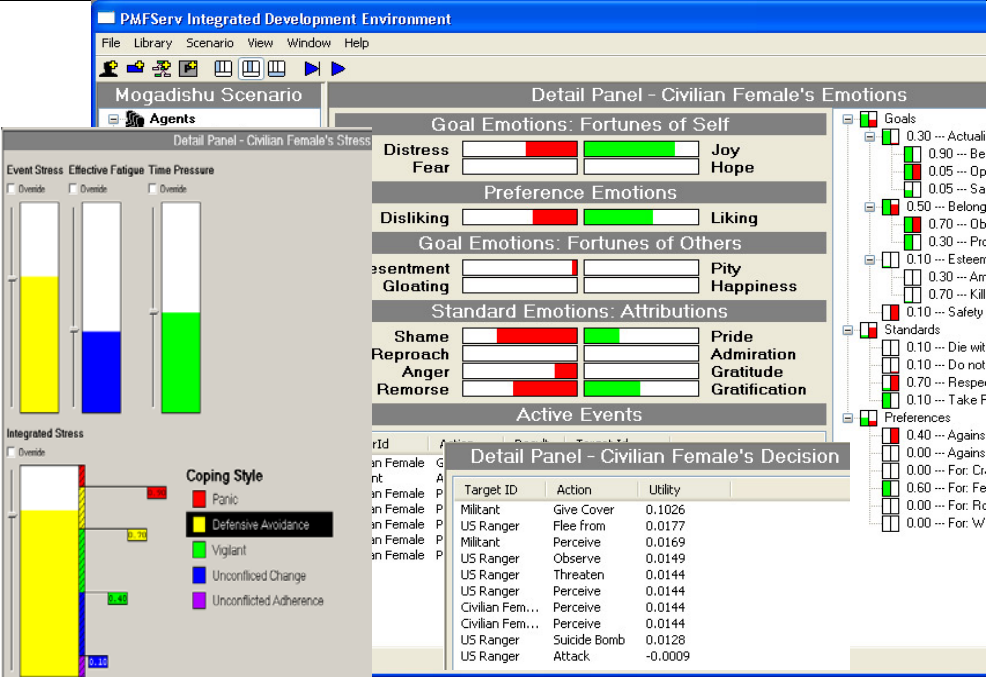
In Figure 9, the player has encountered a member of the Habr Gidr clan's militia. He quickly summons two Bhourka-clad females to surround him and begins shooting from behind them. We do not model verbal utterances in this version of PMFserv, but permit

18

agents to issue software commands to other agents. PMFserv agents are free to obey such commands or not; however, in this case the Somalian females' GSP trees and relationship

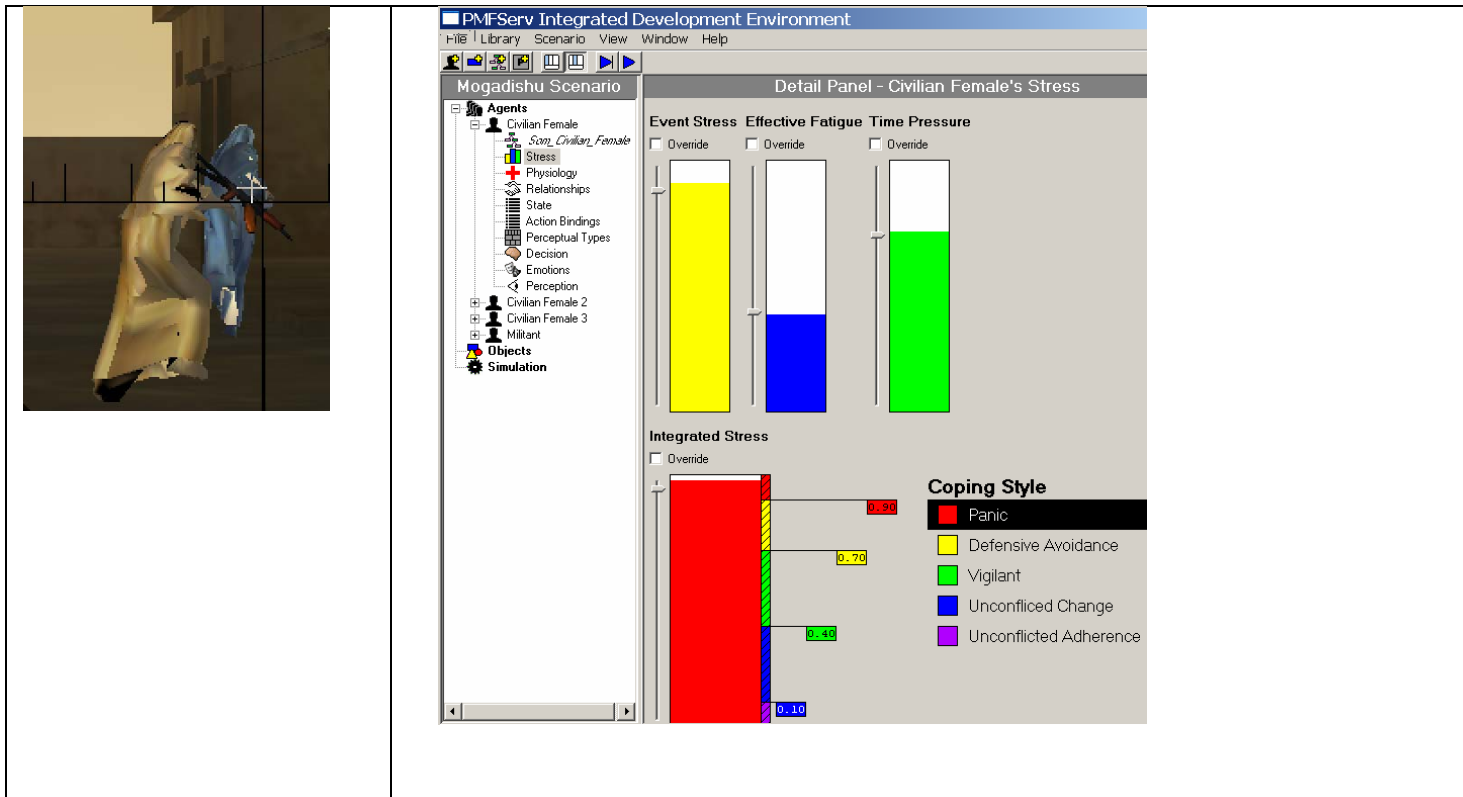| Somalian Female Acting As Shield	Visual Interface to Each Agent's PMFs (Neutral female shield in Coping Mode: Defensive Avoidance, Emotions: Mixed, State: SUBMIT to being a shield)

Figure 9. Somalian Women Acting as Shields, Then Fleeing After the Militiaman is Killed

matrices lead them to find this request to be the highest utility available action. In their relationships they view themselves largely as objects and property of the males in the clan. In their GSP trees we gave them low weight for safety, high weight for belonging to their group, and high weight for revenge on those who do not respect others. Since the Americans, in their view, have committed many past events of disrespect (e.g., helicopters' downwashes that make their bhourkas fly up and that make them drop their infants, attacks on their clan members, etc.), acting as a shield is their way to participate in the revenge action, and it affords them positive emotional activations.

The right side of Figure 9 shows one of the female's various PMF activations. When she is summoned to act as a shield, she exhibits both positive and negative emotions. On the positive side, she gets some *joy* and *pride* from participating, plus she is *happy for* the militia member and *gloating* about the Ranger's predicament. On the negative side, she has *distress* and *shame* from the current situation but also probably from her recent memory of past Ranger violations (stored as the perceptual type she recalls when viewing the Ranger). She feels *dislike* and *resents* the Ranger's presence, and she has *pity* for herself and the rest of her clan. On balance, there is a lot of noise and event stress shown leading to integrated stress or coping mode at the *defensive avoidance* level. As a result, her decision is to submit to the request (to be a shield) which she computed as the highest utility choice. In the lower row of Figure 9, the militia member has just been killed by the Ranger despite his shooting from behind the women. This has caused the female to reach

20

highest *event stress* and the resulting *hypervigilant or panic mode*. Her utility calculations are limited to either cowering in place or drop everything and flee. She chooses the latter, but here is a spot where the legacy simulator interferes. Some bit of code in Unreal (probably written to enhance player enjoyment) always places a loose gun into the hands of a nearby agent. Even though PMFserv has her choose to drop everything and *flee* (and recall she never held the gun to begin with), Unreal has erroneously altered this -- in our application this is a 'bug' we can't remove. Ironically, in the real world of Black Hawk Down there were many instances of women (and children) retrieving weapons of fallen militia. So this is a bug that works in our favor.

5) Results Analysis and Discussion

Part I of this article presented a unified architecture for behavior and a computer implementation known as PMFserv. PMFserv is a parameter-rich system that straddles physiology/stress, affect/personality, social/cultural, and cognitive variables that influence perception and coping behavior. This is a complex system with great power. The current article, Part II, serves as an existence proof that this power can be harnessed and implemented to enhance agent realism and to produce culturally 'interesting' results.

Specifically, we were given a test scenario for this existence proof. To pass that test we did not have to recreate all of the book of *Black Hawk Down*, only the behaviors described in Section 1.1. To pass this test, we authored four archetypes from the culture in question (Somalian militiaman, male and female civilians, and Habr Gidr clan leaders), and used these to populate the virtual Bakarra Market with about two dozen agents. Section 4 presented some of the results so the readers can judge for themselves if these recreate behaviors of their real-life counterparts.

In the end, the reader cannot observe all the detail and nuances of the gamebot behaviors. One turns to judges for that purpose. In our case, the acid test was if the sponsor and their technical representative accepted the results as satisfactory. The sponsor indicated the results were excellent and the technical representative has provided positive reports. Here are several excerpts from the DMSO technical representative's after-action report on the Mogadishu re-creation efforts: *"the affordance-based perceptual subsystem introduced a revolutionary new way to model and simulate early, middle, and late perceptual processes, a research agenda that began at MIT in the mid 1980s with insect-like robots and real-time autonomous video game playing intelligent systems, including Pengi and Sonja* (Agre & Chapman, 1987). *In some respects this work follows that tradition, but the technical and scientific advances are truly significant compared to the original work." "The integrated architecture [PMFserv, ICT, and Unreal] evolved to push the state of the art of intelligent non-playing characters or synthetic agents that could eventually transition to other applications outside of the first-person shooter genre." "In sum, the need for standardization and interoperability of HBRs is becoming an exceedingly critical issue in DoD modeling and simulation efforts and in the gaming and entertainment industries… The final products... at ICT and UPenn were a success in terms of both basic and applied research."* (Toth, 2004)

In another testimonial, according to the prime (ICT) responsible for integrating our PMFserv with the other two Human Behavior Models (Soarbots and AI-Implant) in the many agents of the scenario:

"*The primary result of this effort is the Mogadishu scenario itself. Unlike the heavily scripted play of most commercial games, this scenario is very dynamic and can play out in a wide variety of different ways. This is primarily due to the autonomy and wide range of behavior supported by the three human behavior models. This scenario demonstrates the key contribution of this research; the integration of three HBMs into a single virtual environment through variations on a common interface architecture.*" (van Lent et al., 2004a)

Finally, in viewing the relative contributions of the three human behavior models, the integrator further stated that PMFserv "*demonstrated a higher degree of fidelity in the key areas of emotion modeling, stress and coping styles than the other two human behavior models explored in the project.*" (van Lent, 2004b). Of course the best tactical military decisionmaking came from the Soarbots, while the AI-Implant bots exhibited the best physics, flocking, and navigating.

Besides successfully completing the overall existence proof, this article also explored the answers to three sets of questions posed in Section 1. The remainder of this section returns to those questions and provides a subsection for each that analyzes the results and discusses any lessons learned.

5.1) Are Literature Models Usable and Useful?
Are models drawn from the literature useful and usable as agent minds? To what degree will they elevate an automaton into a realistic agent? Under what conditions do these models help agents pass (fail) correspondence tests?

As readers could observe, using models of physiology derived from first principles, PMFserv guided the gamebots to defendable levels of fatigue, adrenaline and Khatt-drug surges, and trauma. Using a respected opinion leader model of stress and coping mode (Janis & Mann, 1977), calibrated to a gun-inured Bakara marketplace, PMFserv governed when agents would panic and flee and when they would broaden their perception and react more deliberately. Using several respected opinion leader models of emotion (Ortony, Clore, & Collins, 1988; Damasio, 1994), combined with a decision theoretic focus (subjected expected utility) and calibrated with Bayesian weights on GSP trees, the gamebots were guided to select from a wide array of potential action choices that were seen as corresponding to the personality-and culture-specific behaviors one expected of the Habr Gidr clan members. Thus, they exhibited behaviors such as but not limited to (1) unarmed looters emboldened by loot and fellow clan members to swarm armed-yet-not-yet-firing Rangers and retreating only as the Rangers turn violent; (2) Somalian females thinking of themselves as objects and acting as cover for a militiaman, but panicking and fleeing if the militia is killed; and (3) civilians who turn combatants and militia who commit suicide bombings when their beliefs are sufficiently violated. None of these behaviors were scripted and locked in. Via the PMFserv stress and emotion guided utility

processes, these behaviors emerged dynamically from the agents as PMF reservoirs are filled and/or emptied and depending on the actions of the player and other Ranger bots.

The approach of culling PMF s from the literature and coupling them into a unified architecture thus works in toto. Some of the pros and cons that we encountered include:

PROS
- The unified approach permits the consideration of elements of the interplay between biological/stress, affective/personal/cultural, social, and cognitive factors upon agent perception and coping behavior. Most behavior observed in PMFserv agents is the result of all these subsystems interoperating. As a result, sometimes surprising synergies arise from this interoperation. A runtime example of synergy is when agents resolve contradictory information stored in diverse PMFs (preserve self, die for cause, do not kill, women are objects, etc.). Another runtime example is that although GSP weights are fixed and no single PMF incorporates learning, agents are highly adaptive. As certain of their needs and desires are satiated, others rise in importance. This leads to emergence of macro-behaviors in crowds, and to other forms of coping.
- There are also design time synergies from the collection. For example, the case study is for modeling humans from another culture. PMFserv originally sought to implement specific PMFs, with the hope that competing PMFs could replace the original set if warranted. However, with existing PMFs in the affect, social, and decision modules, we were able to implement many aspects of Eidelson's individual's value systems, and Feltovich's cultural identities/norms. This research made us aware that PMFserv might be able to implement cultural and personality models atop existing PMFs, a point we prove formally for a personality instrument in Silverman and Bharathy (2005)
- Within the Biology Module, via physiology tanks/pumps/valves we were able to accommodate literature on PMFs for factors of direct relevance to the scenario such as multiple wounds failing to kill the enemy, Khatt weed affecting performance, etc. In particular, Bharathy (2003a, 2003b) developed and calibrated tanks for trauma from various types of weapons, impact of stimulants such as Khatt,, and fatigue and exertion. These lead to such behaviors having realism in the eyes of the simulation observers.
- The social model supports tanks/valves/pumps on a number of relationship parameters, as well as alignment and group/role scales suggested by the literature (see Part I). In the case study, reports about different groups, roles that archetypes played, and relationship dynamics were successfully accommodated in these structures. As one example, we were readily able to denote women as objects in the eyes of the militia, and women willing to be cooperative when requested to do things counter to some of their own 'instincts'.

CONS
- PMFserv makes an attempt to encapsulate PMFs from the literature in the effort to help users to calibrate and test those PMFs in isolation from the collection. Since most behavior observed in PMFserv agents is the result of many subsystems and PMFs interoperating, this alters the validity of any given PMF at runtime. Many of our inter-operation heuristics are themselves available for study, and most PMFs include viewers so one can see the impact each is contributing to the overall behavior. Nevertheless,

finding explanations for behavior can be a time-consuming effort requiring significant familiarity with the PMFs.

- A scenario may suggest a PMF of interest, but the science of that might be weak – no first principles. For example, we don't yet know what action tendencies the Hofstede cultural dimensions suggest (though that is an example where the missing science is rapidly being filled in).
- Our approach is to study the interactions between many PMFs and modules. Putting this together calls for accuracy rather than precision. For us, 'better is the enemy of good enough' and we make use of linear implementations and first approximations of all PMFs. Future researchers might very well like to alter our implementations, add non-linearities (e.g., bio-rhythms), and drill into shadings of causality behind the behaviors. All PMFs have GUI override switches, and the object oriented encapsulated implementation supports plugin of replacement PMFs. So this is possible.
- If your goal is to build and operate a single and simple scenario, PMFserv has too much power and too little learnability. In each of the modules, we have tried to build the subsystem from lower level PMFs that can be calibrated. This has led us to what some might view as a complex, parameter-rich approach. Clearly, not all parameters are needed for every scenario we seek to simulate. By the same token, however, there is power in the richness of this approach, and we believe this framework has potential to support many kinds of studies and scenarios.

5.2) Software Interchange Lessons
Is the legacy simulator community (military and entertainment) ready and able to accept such plug-in models for updating the minds of bots that already exist in their software? If not, what obstacles exist and what fixes appear warranted?

From the PMFserv perspective, we interfaced with Unreal Tournament via the MS COM interchange method. This interchange protocol performed quite well in practice and did not lead to latency of note in the responses of the bots. What follows is a summary of the observed pros and cons of this approach.

PROS of the Interchange Architecture
- Uses a standardized software approach that's widely available on all PCs
- Microsoft's COM layer is straightforward, well documented, and rapid to implement
- Runtime performance was excellent – no noticeable latency between events and responses -- for up to about six to eight bots in view at once (this is roughly the same performance as UTI itself).

CONS of the Interchange Architecture
- COM is a Microsoft artifact, and not a universal standard
- Limits portability to platforms using Windows
- COM approach doesn't solve many interchange issues, but pushes most of the interchange responsibility onto other layers
- Since there are no naming conventions or translation standards in general for human behavior models, the resulting Custom Unreal Script was difficult to create and grew to about 1,000 lines of code, code that is not itself very reusable.

Due to time and budget constraints, most of the custom UnrealScript had to be dedicated to nuances of this interchange environment and more specifically to this exact scenario. Given a few more such interchanges one might observe some useful patterns and conventions might emerge that would further help the field of human behavior model interchange. Certainly that is a worthy goal and a trend that should be encouraged in the field as more M&S environments attempt to benefit from existing and complementary types of human behavior models.

An interesting commentary about the state of the art in HBM interchange was voiced by DMSO. They recently convened a workshop of human behavior modelers to explore if the field is mature enough yet to start to adopt standards that will help with many of the issues such as those enumerated above (Bjorkman, Barry, & Tyler, 2001). Some of the findings of that workshop include that (1) the field lacks a simple taxonomy or thesaurus of terms and of names items to be modeled thereby making communications more difficult between modeling groups; (2) there are no agreed upon ways to represent human performance data that models might be built from; (3) processes for capturing and representing task and behavioral knowledge fundamentally differ across groups and its not easy to translate between them; (4) there are no standard ways to measure one modeling technique against another, nor are their ways to convert from one to another; and (5) more affordability appears tied to making advances in these topics as well as to increasing the reusability of existing human behavior models across simulators. Our findings in this study are compatible with all of these, and the fifth topic is a good point to transition to the final set of questions we investigated here.

5.3) Composability and Knowledge Engineering Lessons

What is needed to improve the composability situation so that digital casts can be created? From a knowledge engineering perspective, how do various methods and approaches impact affordability?

In addition to documenting the 'scenario test', this article uses the Mogadishu case study to illustrate a methodology (4 stages, 5Ps, 6 steps, etc.) for using PMFserv to build and operate digital casts that enhance simulator agent realism. This is a methodology we have used on several similar studies and that our developer community is beginning to have success with as well. It is not perfect and it's a methodology we continue to refine as we expand it and learn more about how people use it (e.g., see Bharathy, 2005).

PROs
- The 4 stage methodology of Figure 1 fits the purpose it was created for, namely to develop simulation scenarios with realistic characters by synthesizing scientific principles and behavior models into PMFserv and embedding them behind legacy simulators.
- The six behavior model authoring steps of Section 3 help PMFserv developers to bridge the gap between anecdotal reports/qualitative literature materials, expert opinion,

behavior model specifications (evidence tables, Bayesian weights), and PMF implementation and tuning. Spreadsheets, spreadsheet macros, and visual programming assist this process. Advanced users also can access underlying code editors.

- GSP trees are a useful way to capture and represent value systems. In some cases where only anecdotal and textual evidence exists, our method uses evidence tables for moving textual statements from sources directly into Bayesian weights. If large data sets are available, one can derive statistical likelihoods, prior odds, etc. with the same procedure. GSP trees are also usable for implementing many types of personality instruments directly as nodes on the trees. In such cases where instruments include profiling methods, then these can be supported as well. We recently did this with Hermann's Political Leader Profiling Instrument (see Silverman and Bharathy, 2005).
- Part I of this article elaborated on how the affordance markup approach reduced the complexity and maintainability of our agents. The present scenario effort involved a significant test of the scalability of this approach that it passed in several ways. We were able to get students to readily understand it, fill in spreadsheets, use the visual editors, and markup the world of objects with their perceptual types, available actions, and afforded results. Thus, it not only passed a usability and workability test, but it also passed the usefulness test for this scenario.

CONS

- To date we have largely used the 4 stage methodology for composition of training applications, and have only recently begun research on analytical uses. At a minimum, extensions are needed in the area of design of simulation experiments, and how to guarantee convergence on robust solutions in complex parameter space.
- The six step process does not presently consider human cognitive biases. We have observed that developers when assembling evidence tables tend to anchor on a single hypothesis to explain behavior and then seek only confirming evidence. Bharathy (2005) is researching how to introduce differential diagnosis and the consideration of alternative competing hypotheses directly into the spreadsheet support approach.
- Our sponsors to date have only been willing to fund rather limited proof of existence tests and correspondence validations. The Mogadishu study is a case in point. To fully understand and trust agent behavior models, a number of validation tests should be supported such as individual PMF tests, further correspondence tests, Turing tests, and competing agent model tests, among others. A suite of software to support regular testing is called for as well.
- As a final lesson learned, substantial effort was necessary to markup the objects with affordances, to cull various relevant sources, and to 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 of them in other simulators.

6) Conclusions and Next Steps
The main purpose of this effort was a case study to explore how to integrate "off the shelf" human behavior models into pre-existing game engines and M&S environments in order to enhance the realism of the characters in different roles. This was accomplished

by building a standard interface to a commercial game engine and tuning an "off the shelf" human behavior model (PMFserv) to populate the scenario.

All the results and judges' statements reported here are encouraging. However, this was but a single test for a relatively small scenario. It remains to apply the PMFserv capability to other tests, and for third parties (not the developers or their students) to try to use it to implement other models from the literature. Clearly, we are only at the beginning of a long process, one that we hope but can't guarantee will open up new vistas for collaboration across disciplines and for improving the realism of and value of agent models and simulation.

The enterprise of human performance simulation is too vast an undertaking for any one provider to have it all. 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. We hope that our research will help contribute to that advance, as summarized in this two-part article.

When and if the field conquers these interoperability 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 behavioral modeling, which will shift from reductive, silo-separated specialties, 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. 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.

References

Abshir (1998). Interview: Mrs. Abshir. *Frontline: Ambush in Mogadishu.* PBS Online.
        <http://www.bs.org/wgbh/pages/frontline/shows/ambush/interviews/abshir.html>.

Agre, P. E., & Chapman, D. (1987). Pengi: An implementation of a theory of activity.
        Proceedings of the Sixth National Conference on Artificial Intelligence, Seattle,
        1987: 196-201.

Anon. (1995). Department of Defense Modeling and Simulation (M&S) Master Plan,
        Under Secretary of Defense for Acquisition and Technology, Washington, D.C.,
        October 1995.

Bharathy, G. (2005). *Modeling the behavior of synthetic human agents in virtual worlds:*
        A systems methodology for carrying out knowledge engineering & modeling of
        cognitively detailed agents. Doctoral dissertation proposal, University of
        Pennsylvania, 2005 (available from the author at bharathy@seas.upenn.edu).

Bharathy, G. (2003a, May). *Physiology Tank Report*, prepared for SYS 508,
        Philadelphia: University of Pennsylvania (available from the author at
        bharathy@seas.upenn.edu).

Bharathy, G., Damghani. B, Kim, E., & Lambert, L. (2003b, May). *Design of Agent-*
        Injury Models, prepared for SYS 508, Philadelphia: University of Pennsylvania
        (available from the author at bharathy@seas.upenn.edu).

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.

Bjorkman, E. A., & Blemberg, P. (2001, Spring). *Review of the Defense Modeling and*
        Simulation Office Human Behavior Program, Simulation Interoperability
        Workshop Paper Number 01S-SIW-080.

Bowden, M. (1999). *Black Hawk Down.* New York: Atlantic Monthly Press.

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

Eidelson, R. J., & Eidelson, J. I. (2003). Dangerous ideas: Five beliefs that propel groups
        toward conflict. *American Psychologist, 58,* 182-192.

Farah, N. (2000). *Yesterday, Tomorrow: Voices from the Somali Diaspora.* New York:
        Cassell.

Feltovich, P., Bradshaw, J. M., Jeffers, R., Suri, N., & Uszok, A. (2004). Social order and
        adaptability in animal and human cultures as an analogue for agent communities:
        Toward a policy-based approach. In *Engineering Societies in the Agents World*
        IV. LNAI 3071. (pp. 21-48). Berlin, Germany: Springer-Verlag.

Finerman, L. E., Prochnow, D. L., Gluck, K. A., & Willis, R. P. (2001, Spring). *A*
        Persistent Federation for Human Behavior Representation Models, Simulation
        Interoperability Workshop Paper Number 01S-SIW-058, Spring 2001.

Fullerton, T., Swain, C., & Hoffman, S. (2004). Game design workshop: Designing,
        prototyping, and playtesting games. New York: CMP Books.

Gibbons, A. S., & Fairweather, P. G. (1998). *Computer-based instruction: design and*
        development. Englewood Cliffs: Educational Technology Publications.

Gilbert, N., & Troitzsch, K. (1999). *Simulation for the social scientist*. Buckingham: Open University Press.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Newbury Park, CA: Sage.

Hussein, I. (1997). *In their own voices: Teenage refugees from Somalia speak out.* New York: Rosen Publishing Group, Inc.

Jacobson I. (1992) Object-oriented software engineering: A use case driven approach. Reading: Addison-Wesley.

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

Lavine, N., Peters, S., Napravnik, L., & Hoagland, D. (2002). An advanced software architecture for behavioral representation within computer generated forces. In *11th Conference on CGF & BR*, Orlando, FL: SISO, May 2002, pp. 169-180.

Lewis, I. M. (1994). *Blood and Bone: The Call of Kinship in Somali Society.* Laurenceville, NJ: Red Sea Press, Inc.

Lombardo, J., Pollack, B., George, M., & Brounstein, A. (2003). *Senior Design Team on Mogadishu (Modeling human behavior within the "Black Hawk Down" scenario).* University of Pennsylvania, Senior Design Project Report, January 2003.

Nelson, H. (Ed.) (1982). *Somalia: Country study*. Washington, DC: United States Government, 1982.

Nisbett, R., Choi, I., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin, 125*(1), 47-63.

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.

Runals, S, (2004). *Transformational capabilities: Effects based planning and operations*. Educational Briefing Slides, Joint Forces Command, SJFHQ(CE)–S/T, available from ca.dtic.mil/doctrine/july04_sjfhq_onaeffects_info.ppt

Saaty, T. L. (1982). Decision making for leaders. New York: Van Nostrand Reinhold.

Schreiber, G.(1999). *Knowledge engineering and management: The CommonKADS methodology.* Cambridge: MIT Press.

Silverman, B., & Bharathy, G. (2005). Modeling cognition and personality in leaders. 2005 Conference on Behavior Representation and Simulation (BRIMS), SISO, May 2005 (pending).

Silverman, B. G., Cornwell, J., O'Brien, K., & Johns, M. (2004). *Human behavior models for agents in simulators and games: Part I – Enabling science with PMFserv.* (Submitted for publication)

Silverman, B.G., O'Brien, K., Cornwell, J.B., (2003, August). *Mathematical theory of software agent behavior and human behavior models (HBMs) to increase the realism of synthetic agents.* ACASA/UPenn, Tech Report, (available from www.acasa.upenn.edu/Final_Tech_Report.doc).

Toth, J. A. (2004). *After-Action Report: Integrating disparate human behavior representations in a first-person shooter game based on the Blackhawk Down incident in Mogadishu.* (IDA Document D-2975). Alexandria, VA: Institute for Defense Analyses.

Toth, J., Graham, N., van Lent, M., 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.

van Lent, M., McAlinden, R., Brobst, P., et al. (2004a). Enhancing the behavioral fidelity of synthetic entities with human behavior models. *13th Conference on Behavior Representation in Modeling and Simulation (BRIMS),* SISO, May 12-15, 2004.

van Lent, M. (2004b). Personal communication, Summer, 2004.

Williams, S., & Kindel, C. (1994). *The component object model: A technical overview.* Redmond, WA: Microsoft Corporation.