View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarlyCommons@Penn

University of Pennsylvania
ScholarlyCommons

Departmental Papers (ESE)

Department of Electrical & Systems Engineering

September 2007

Modeling Factions for 'Effects Based Operations': Part I Leader and Follower Behaviors

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

Gnana K. Bharathy
University of Pennsylvania

Benjamin Nye
University of Pennsylvania

Roy J. Eidelson
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/ese_papers

Recommended Citation

Barry G. Silverman, Gnana K. Bharathy, Benjamin Nye, and Roy J. Eidelson, "Modeling Factions for 'Effects Based Operations': Part I Leader and Follower Behaviors", . September 2007.

Postprint version. Published in *Computational & Mathematical Organization Theory*, Volume 13, Issue 3, September 2007, 30 pages. Publisher URL: http://dx.doi.org/10.1007/s10588-007-9017-8

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/ese_papers/294
For more information, please contact repository@pobox.upenn.edu.

Modeling Factions for 'Effects Based Operations': Part I Leader and Follower Behaviors

Abstract
This paper presents a synthetic approach for generating role playing simulation games intended to support analysts (and trainees) interested in testing alternative competing courses of action (operations) and discovering what effects they are likely to precipitate in potential ethno-political conflict situations. Simulated leaders and followers capable of playing these games are implemented in a cognitive modeling framework, called PMFserv, which covers value systems, personality and cultural factors, emotions, relationships, perception, stress/coping style and decision making. Of direct interest, as Sect. 1.1 explains, is mathematical representation and synthesis of best-of-breed behavioral science models within this framework to reduce dimensionality and to improve the realism and internal validity of the agent implementations. Sections 2 and 3 present this for leader profiling instruments and group membership decision-making, respectively. Section 4 serves as an existence proof that the framework has generated several training and analysis tools, and Sect. 5 concludes with lessons learned. Part II turns to the question of assessment of the synthesis and its usage in course of action studies.

Keywords
leaders and followers, strategy games, personality and culture, agent-based simulation

Comments
Postprint version. Published in *Computational & Mathematical Organization Theory*, Volume 13, Issue 3, September 2007, 30 pages. Publisher URL: http://dx.doi.org/10.1007/s10588-007-9017-8

This journal article is available at ScholarlyCommons: http://repository.upenn.edu/ese_papers/294

Modeling Factions for 'Effects Based Operations':
Part I –Leaders and Followers

Barry G. Silverman, Gnana Bharathy, Benjamin Nye, Roy J. Eidelson
Electrical and Systems Engineering Dept.
University of Pennsylvania, Philadelphia, PA 19104-6315
basil@seas.upenn.edu

"If sociological game theory is not to end up as an artificial exercise, ..., it is absolutely essential that the beliefs, ideas and experiences of the actors themselves are moved onto center stage"
R. Swedberg (2001), p. 325

This paper presents a synthetic approach for generating role playing simulation games intended to support analysts (and trainees) interested in testing alternative competing courses of action (operations) and discovering what effects they are likely to precipitate in potential ethno-political conflict situations. Simulated leaders and followers capable of playing these games are implemented in a cognitive modeling framework, called PMFserv, which covers value systems, personality and cultural factors, emotions, relationships, perception, stress/coping style and decision making. Of direct interest, as Section 1.1 explains, is mathematical representation and synthesis of best-of-breed behavioral science models within this framework to reduce dimensionality and to improve the realism and internal validity of the agent implementations. Sections 2 and 3 present this for leader profiling instruments and group membership decision-making, respectively. Section 4 serves as an existence proof that the framework has generated several training and analysis tools, and Section 5 concludes with lessons learned. Part II turns to the question of assessment of the synthesis and its usage in course of action studies.

Keywords: Leaders and followers; strategy games; personality and culture, agent-based simulation

1) Introduction and Purpose

Gaming and simulation of socio-cultural groups is a newly evolving field, motivated by the need to better understand how leaders and followers behave, what motivates them, how dangerous ideas spread, and how they might be influenced to cooperate, mitigate conflicts, and benefit the overall good. Green (2002) studies the array of methods for forecasting conflict and shows that predictions are significantly improved when subjects first participate in role playing games about the issues at stake. Hence, one aim of this research is to isolate the components needed for a generic role playing game to be used to rapidly

mock up a class of conflicts commonly encountered in today's world. In other words, create a widely applicable game generator. Since it is often impossible to find humans to play all the roles of such games, or to play out all the possible scenarios, a second aim is to create plausible models of leaders and followers based on first principles about what makes them tick and so they may play some of the roles in the game. If these cognitive agents are realistic, they can help trainees and analysts explore the range of their possible actions under varieties of conditions, thereby helping others to see more clearly how to influence them and elicit their cooperation. A related assumption based on evidence from video- and multi-player online-games, is that if the agents have sufficient realism, that should further motivate players (trainees) to be engaged and immersed in role playing games or online interactive scenarios. The benefits of these first two goals lie in forcing trainees and analysts to arrive at explanations of what is going on in a situation.

A 'catch-22' of these two goals is that, agent based simulation games will be more valuable the more they can be imbued with realistic leader and follower behaviors, while the social sciences that can reliably contribute to this undertaking are made up of many fragmented and narrow specialties, and few of their ' models' have computational implementations. The third aim of this research is thus to improve the science by synthesizing relevant first principles and best of breed social science models so they can be tested in agent-based games, and by that to cause the limitations of these models to be exposed and improved. What is known in diverse fields such as sociology and political science, psychological and cognitive modeling, anthropologic/culture modeling, epistemology and the spread of ideas, and personality profiling that might help one to construct more realistic models of groups, in general, and their leaders and followers, in particular? Is an agent-based synthesis and implementation feasible?

We turn to a review of some of the literature in Section 1.2, after which Sections 2 and 3 examine best of breed approaches to modeling leaders and followers and how they interact in the game framework we have assembled. In Section 4, we summarize some of the games we have authored and discuss how they form an existence proof that such a framework is feasible. Finally, Section 5 discusses what has been learned in striving for the 3 aims, and a research agenda for improving the field of socio-cultural games and agent based leader-follower modeling and simulation. First, however, we pause to examine the basic socio-culture game theory of interest here.

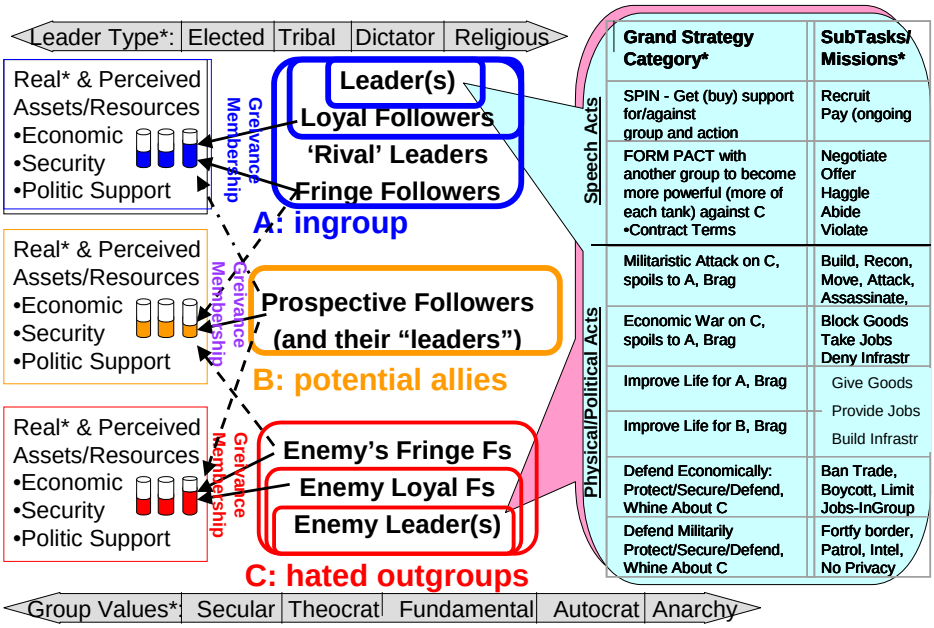


Figure 1 – Overview of Leader-Follower Game Theory (* - editable list)

1.1) Leader-Follower Game Theory

How can an analyst or trainee devise policies that will influence groups for the collective good? And what must a socio-cultural game generator encompass? Figure 1, explained below, attempts to portray a fairly universal class of leader-follower game that groups often find themselves in and that are worthy of simulation studies. This could be for competing groups in a crowd, in an organization, in a region or nation, or even between nations. Analysts would need an appropriate suite of editors and a generator, to help them rapidly mock up such conflict scenarios and analyze what outcomes arise from different courses of action/policies.

Specifically, the socio-cultural game centers on agents who belong to one or more groups and their affinities to the norms, sacred values, and inter-relational practices (e.g., language, gestures, social rituals) of those groups. Let us suppose there are N groups in the region of interest, where each group has a leader archetype and two follower archetypes (loyalists & fringe members). We will say more about archetypes shortly, and there can certainly be multiple leaders and followers, but we stick in this discussion to the smallest subset that still allows one to consider beliefs and affinities of members and their migration to more or less radical positions. There is an editable list of norms/value systems from which each group's identity is drawn. The range across the base of Figure 1 shows an example of a political spectrum for such a list, but these could just as easily be different parties in a common political system, diverse clans of a tribe, different groups at a crowd event, and so on. Each entry on this list contains a set of properties and conditions that define the group, its practices, and entry/egress

stipulations. The authority of the leader in each group is also indicated by a similarly edited list depicted illustratively across the top of Figure 1.

The vast majority of conflicts throughout history ultimately center around the control of resources available to a group and its members. The resources of each group are illustrated along the left side of Figure 1 and are summarized for brevity into three tanks that serve as barometers of the health of that aspect of the group's assets – (1) economy and goods available to the members (jobs, money, foodstuffs, training, healthcare etc.); (2) rule of law applied in the group as well as level and type of security available to impose will on other groups; and (3) popularity and support for the leadership as voted by its members. Querying a tank in a culture game will return current tank level and the history of transactions or flows of resources (in/out), who committed that transaction, when, and why.

To start a game, there are initial alignments coded manually, though these will evolve dynamically as play unfolds. Specifically, each group leader, in turn, examines the group alignments and notices Loyal Ingroup (A), Resistant Outgroup (C), and those "undecideds" in middle (B) who might be turned into allies. Also, if there are other groups, they are examined to determine how they might be enlisted to help influence or defend against the out-group and whatever alliance it may have formed. Followers' actions are to support their leader's choices or to migrate toward another group they believe better serves their personal value system. Actions available to Leader of A are listed in the table on the right side of Figure 1 as either speech acts (spin/motivate, threaten, form pact, brag) or more physical/political acts. Of the latter, there are 6 categories of strategic actions. The middle two tend to be used most heavily by stable, peaceful groups for internal growth and development. The upper two are economic and militaristic enterprises and campaigns taken against other groups, while the lower two categories of actions are defensive ones intended to barricade, block, stymie the inroads of wouldbe attackers. The right hand column of the action table lists examples of specific actions under each of these categories – the exact list will shift depending on whether the game is for a population, organizational, or small group scenario. In any case, these actions require the spending of resources in the tanks, with proceeds going to fill other tanks. Thus the culture game is also a resource allocation problem. Leaders who choose successful policies will remain in power, provide benefits for their followers, and ward off attackers. Analysts and trainees interacting with this game will have similar constraints to their policies and action choices.

The military analyst will recognize this description as falling into a class of games known as effects based operations (EBO) where the operations one undertakes are courses of action that tend to fall into one of the following choices: Diplomatic, Informational, Military, and/or Economic (DIME). Similarly, the outcomes of operations may show up as one of more of Political, Military, Economic, Social, Informational, or Infrastructure (PMESII) categories of effects. In the EBO world, one chooses PMESII

effects that are desired and attempts to study which alternative DIME courses of action will achieve them or lead to potentially unwanted outcomes.

The lead author spent much of 2004 assembling a paper-based version of Figure 1 as a role playing diplomacy game and play-testing it with analysts: Silverman, Rees et al (2005). The goal of the game is to help players to experience what the actual leaders are going through, and thereby to broaden and deepen their understanding, help with idea generation, and sensitize them to nuances of influencing leaders in a given scenario. In zero sum games, what one spends on actions affecting one area of welfare, security, or populace effects what one has to allocate to other areas. The mechanics of the game place the player at the center of the action and play involves setting objectives, figuring out campaigns, forming alliances when convenient, backstabbing when necessary. The idea here was to make the game immersive and engaging, and to date hundreds of players have participated in multi-hour sessions that they were unwilling to terminate. All this game-play also gave us a rich source of data to help guide the construction of agents who can serve as synthetic opponents, allies, followers, and the like.

This is in the genre of the Diplomacy or Risk board games, though unlike Diplomacy, its rapidly reconfigurable to any world conflict scenario. A streamlined version of the paper-based game has been turned into a boardgame called BigWig© to appear at toy stores in early 2007 and aimed at being played to conclusion within an hour (it is thus intended to serve as an intro to the diplomatic strategy genre for new players). Section 4 will return to several computer implementations of this game that have been built and fielded to date.

1.2) Computational Theories of Agents

To our knowledge there are no other agent based (or other) approaches that come close to embracing the breadth and depth in Figure 1, though there are a number of contributions to various slices of that picture. Here we mention a few of these in order to illustrate the range of issues one must model to implement Figure 1. We begin with theories and implementations that are computer based and then proceed to the behaviorally inspired ones.

Game Theory - As a first sub-community, game theory itself offers some useful notions, though for the most part it is a rich theory only when there are very few adversaries, limited action choices, simple payoffs, and a fixed game. None of these constraints fit modern world conflicts and thus game theory alone is incapable of handling the details of Figure 1. Armstrong (2002) and Green (2002) show game theory is no better than chance at predicting conflict in real world case studies. Camerer (2003) makes similar statements about game theory in the absence of realistic theories of economic behavior. He goes on to describe the recent emergence of 'behavioral game theory' which focuses on what players actually do and how it expands analytic theory by adding emotions, errors, biases, and the like. Simari & Parsons

(2004) ran an experiment comparing the convergence rates of (1) approximations and relaxations of prescriptive approaches (i.e., POMDP) and (2) "descriptive approaches" based on how humans tend to make decisions. In simplistic games the prescriptive approximations do better, but as games grow larger and more complex the descriptive approach is preferred and will provide closer convergence and faster performance. Part II of this paper returns to analytical game theory and how the models of this article (Part I) can inform and potentially improve the performance of game theory for ethno-political topics.

Communicative and Plan Recognizing Agents - A related game-agent sub-community involves straddling the artificial intelligence and epistemological literatures in what is at times called nested intentionality modeling. In this type of approach one finds theories and implementations of agents attempting to move beyond just observing agents' actions to also include the modeling of intentionality of the other agents in the game.This is a nascent community, although some progress has resulted from the earliest days where it was claimed that the problem was NP-complete: Kautz & Allen (1986). Some researchers have advanced the field via formal methods such as POMDP, Distributed Bayesian Networks, etc. such that their plan recognizers tend to operate in polynomial space: Geib (2004). However, these tend to be for rather simple problems, such as 2 player games with minimal action choices. Other researchers have pushed the complexity to the O(linear) level primarily by focusing on descriptive heuristics relevant to the domain in which they work, rather than trying to apply complex, principled formalisms: e.g., see Kaminka & Avrahami (2004). It is this latter type of paradigm that we pursue in our own intention modeling. The large agent communication literature also provides useful methods for speech act performative manipulation, inter-agent communication, and negotiation management that we omit from the current discussion, but which benefits and informs our work as well.

Strategic RPG Videogames - For the most part, AI in videogames is devoted to the use of finite state machines for low-level functions such as navigating, path finding, and personal combat tactics. A small portion of this community focuses on what might be called "Leader AI" mostly for strategy games, but also role playing games and a few massive multiplayer online games. Most often these games have the player try to run a country or constituency in some historical era or in a fantasy or futuristic alien world. The leader AIs tend to be the opponent leaders and they fairly uniformly rely on simple rules buried in scripts to govern the branching of agent behaviors. The AI is there mostly to entertain the player, and rules are kept simple and pre-scripted, often with an easy to beat opponent for new players and a harder level to challenge better players (accomplished by making the AI omniscient). These AIs tend to be difficult to communicate with except in simple ways, are hard to make alliances with at anything but the highest levels, are incapable of carrying out coordinated efforts, and often are inconsistent: e.g., see Woodcock (2002).

Human Behavior Models - An important learning objective of our research program is to directly model and explore the underlying behaviors and motivations of leaders and followers. For these purposes, we turn to a final agent sub-community -- that of the human behavior modelers. Unlike game AI, human modeling need not focus on navigating and battling around spaces, but instead invests in detailed modeling and simulation of individual agent's decision and cognitive processes, and worries about accuracy and validity of the resulting behaviors. This community eschews the rationalistic-normative paradigm of game theory, and instead often embraces an array of descriptive, naturalistic, or heuristic approaches to modeling human cognition and reasoning: e.g., see Pew & Mavor (1998). As cited above, descriptive/cognitive approaches: (1) hold a better chance of converging in coalition games; (2) are more tractable for plan recognition/intentionality modeling; (3) offer improved mechanisms for the perception and reasoning about aspects of other agent behavior such as trust, relationships, and group membership choices; and (4) hold the promise to help agents move beyond surface features and explain their choices and behaviors. Despite these arguments for richer human behavior models, the human behavior modeling community does not hold a well-developed approach to all these issues. There are some widely used cognitive models (e.g., SOAR, ACT-R) that tend to be quite adept at logical processes of reasoning and learning but which ignore many other components of behavior and cognition. In contrast, PMFserv complements such models of cognition with a wide range of implemented best-of-breed human performance moderator functions (PMFs) such as: for physiologic impacts (e.g., exertion, fatigue, injuries, stimulants, noise); for sources of stress like event stress or time pressure; and with models of emotion and affect to modulate expected utility formulations and to help account for cultural, personality, and motivational factors that impact upon leader-follower relations, Silverman (2005). PMFserv also manages social relationship parameters and thus macro-behavior (e.g., in collectives or crowds of PMFserv agents) emerges from individuals' interactions and micro-decisions. As a result, PMFserv is the agent framework adopted for our leader-follower AI. The human behavior modeling community is relatively new to the idea of modeling individual differences drawn from validated personality/culture models and instruments. That is the portion of PMFserv we focus on here, as the next two sections explain for leaders and followers, respectively. Readers interested in more details about PMFserv's many other modules (emotion, perception, physiology, stress, etc.) are referred to Silverman et al (2006a). Those details are omitted here.

1.3) Dimensionality Reduction and the Systems Approach

DIME-PMESII game and agent models such as proposed here quickly run into the 'curse of dimensionality' where the number of parameters can explode far beyond what any training dataset can support (de Marchi, 2005). One way to reduce dimensionality is to simplify the game and agents, as is often done in game

theory when applied to conflict. However, this runs the risk of solving problems that bear little resemblance to the real world. A better way to reduce dimensionality is the systems approach that focuses on breaking a system into components (sub-systems) with encapsulated functionality, preserving inter-relationships between components, and applying domain theories/knowledge to keep realism in the components to the extent possible. In such a hierarchically organized system, the relationship between parameters itself reflects additional information or knowledge. Lower levels (say leaves) of parameters can be aggregated into higher nodes, where these fewer, higher level parameters are equally good predictors. This is the approach taken here, and it can be applied recursively to any component. For example, the leader-follower game of Section 1.1 encapsulates and isolates parts of the game into components that are interlinked so they can be modeled at varying resolution without affecting how the collection of components interact. An example of this is the resource tanks that leader agents perceive as a stack of poker chips that grows or falls. We actually plug in finer resolution models for any given tank without affecting overall system performance (e.g, economy models, population support models, etc.). The agent perceptions of these resource tanks, is further abstracted to 3 settings (low, medium, high) when assessing power and vulnerability in the world. Thus agents have a minimum of possible parameters to focus on even though the models of these resources might be hierarchical and complicated.

A different type of example concerns the parameters internal to a given agent. A naïve reader will observe the following sections introducing hundreds of internal psycho-social parameters (values, emotions, stress, relations, etc.). Once again, the systems approach prevails and these are in components with inter-relations between them nicely externalized. Further, a standard way to reduce dimensionality problems is to use domain theories and knowledge (de Marchi, 2005). That is the point of best of breed models and first principles. Such models reduce the dimensionality to the traits and factors they require, and where these are applied, we can use training datasets, fill in the traits and factors of archetypical as well as real characters, conduct validation tests, and treat these parameters as no longer existent. That is they are no longer independent variables clouding the larger DIME-PMESII analyses, but are 'swept out of the way' by first principles and training data before DIME-PMESII studies even begin. Thus the systems approach provides a dual benefit of synthesizing the social science models into a wholism at the same time that it uses them as domain knowledge to remove these parameters from the frame for DIME-PMESII analysis. Even if we hypothesize two versions of a given leader (e.g., one more benevolent the other autocratic), the hundreds of parameters inside them are reduced down to just those two. This is not to belittle the effort that goes into producing such characters for DIME-PMESII studies. In fact that is why it requires two parts of this article – Part I on the syntheses that ultimately lead to dimension reduction, and Part II on the DIME-PMESII studies that benefit from having highly realistic characters.

Some authors argue against dimensionality reduction since it can lead to a loss of information and resulting distortion: e.g., see Cilliers (1998). In this case, we disagree since one can retrieve information when it is needed. The dashboard of a car does not show the driver the air-fuel ratios in the carburetor or

firing chamber (or the 1,000s of other parameters in a car), but the mechanic can display those on his diagnostic panel when he needs them. This leads to no loss of performance or distortion to the driver. Likewise, other authors may object to the idea that components with explicit inter-relations sounds more like a complicated system, than a complex one: e.g., Buckley (1967), Cilliers(1998). However, a PMFServ system consists of agents and objects in the world, the states of which determine the full model state. Agents generate decisions according to their individual perception of this mutually shared world, as well as internal components such as physiology and emotional state which are not directly known to other agents. This results in a system which is highly non-linear and connected, with each component in the world (internal mental models and physical entities) updated by what could be expressed as partially coupled discrete-time recursions.

2) Theories on the Personality and Values of Leaders

An important source of ideas for the modeling of leaders comes from the fields of political psychology and strategic leadership: eg, see Canella & Monroe (1997). While these are not implemented or operationalized agent models, they can be a source for creating agent frameworks with greater degrees of realism. In terms of leader theories that might impact the leader-follower game (Aim 2), Chemers (1997) indicates that while the leadership literature often appears fragmented and contradictory, it can be summarized in terms of four dominant theories of leadership: Leader Trait Theory, Transaction Theory, Transformational Leader Theory, and Situational Contingency Theory. Canella & Monroe offer a similar taxonomy with a few added theories. All these leadership theories seek to stipulate the most effective leadership style and behaviors that a leader must perform well in order to be effective in a given situation. Our current LeaderSim agents include implementations of some portions of the situational, transformational and transactional features as will be outlined shortly. However, these theories ignore leader to leader relations, plus they tend to be normative and prescriptive. From our viewpoint, we need to embellish this with a theory that focuses upon inter-leader dynamics (such as Uhl Bien and McKelvey (in press) begin to theorize about) and that permits us to model leaders as they are (not as a prescription would prefer). As a result, we turn now to a descriptive theory of leader style, one that is measurable and can be fully implemented in our agent-based framework.

After two decades of studying over 122 national leaders including presidents, prime minister, kings, and dictators, Hermann (1999), has uncovered a set of leadership styles that appear to influence how leaders interact with constituents, advisers, or other leaders. Knowledge about how leaders react to constraints, process information, and are motivated to deal with their political environment provides us with data on their leadership style. Hermann determined that the traits in Table 1 are particularly useful in assessing leadership style.

In Hermann's profiling method, each trait is assessed through content analysis of leaders' interview responses as well as or other secondary sources of information. Hermann's research also has developed methods to assess leadership at a distance, based mostly on the public statements of leaders. While both prepared speeches and statements from interviews are considered, the latter is given preference for its spontaneity. The data is collected from diverse sources, usually as many as 50 interviews, analyzed or content coded, and then a profile can be developed. These are then compared with the baseline scores developed for the database of leader scores. Hermann (1999) has developed mean scores on each of the seven traits. A leader is considered to have high score on a trait, if he or she is one standard deviation above the average score for all leaders on that trait.

Table 1 – The Seven Traits of the Hermann Leadership Style Profile

1. Belief that one can control what happens	Combination of the two attributes (1) and (2) determines whether the leader will challenge or respect the constraints.
2. Need for power/influence	
3. Concept complexity (IQ)	Combination of the two attributes (3) and (4) determines how open a leader will be to information.
4. Self-confidence	
5. Task Vs Relationship Focus	a continuum between two poles: o Moving the group toward completion of a task, and o Maintaining group spirit and morale (building relationships).
6. An individual's general distrust or suspiciousness of others	The leader's outlook about the world and problems largely determines the confrontational attitude of the country, likelihood of taking initiatives and engaging in sanctions. The extent of their in-group bias and general distrust of others is driven by: o perceived threats or problems in the world, or o perceived opportunities to form cooperative relationships.
7. The intensity with which a person holds an in-group bias.	

In our LeaderSim personality model, we adopt Hermann's traits (Table 1) with the following changes:

- simplified traits 3 and 4 by using Openness-to-Information directly rather than as a combination of conceptual complexity and self confidence.
- After discussions with Sticha et al. (2001), we added one further trait, namely Protocol Focus vs. Substance Focus as a continuum to describe the leader's penchant for protocols (e.g., state visits or speech acts such as religious blessings) as opposed to taking any concrete actions.

Our LeaderSim implementation includes a full set of screens and sliders for quickly tuning these parameters for each leader in a conflict scenario. Due to space limits we omit presenting these here, but interested readers should consult Silverman, Rees, et al. (2005). Using our Hermann implementation

moves us toward Aim 2 (best of breed models) as one can populate a game with real leader profiles provided the profiling were done properly. Thus, for example, one could determine which leaders tend to be deceitful vs. honest. Specifically, the leader with low belief in control (trait 1) but high need for power (trait 2) tends toward deceit, while the leader with high trait 1 and high trait 2 tends toward accountability and high credibility. Likewise, the same could be done for the other traits (and our new trait of protocol vs. substance), as we will attempt to demonstrate. Any implementation of a best of breed, paper-based model runs into the catch-22 outlined in Aim 3, and the need to try and enhance that model by integrating it with others and making it executable by autonomous agents. Hermann gives no guidance on how an agent computes if it is achieving its need for power (trait 2). We turn in the next section to a discussion of some of the mathematics needed to complete her model and attempt to interpret it faithfully, yet provide an implementation. Subsequent sections examine our interpretation in terms of face validity, Turing tests, correspondence tests, and sensitivity analyses.

Some readers may think a trait model within an agent leads to a tautology of the form: "We classify people into having levels of traits A, B, C and this causes them to tend toward set of actions X. Therefore these agents will always do X. No surprise is possible." This view would be a misunderstanding of what has been presented here. Each agent is imbued with many competing traits and often faces conflicting needs (e.g., hunger and humanitarianism, survival and sacrifice for group, etc.). Social contexts provide activations and situations that drive one or the other of the traits and needs differently and sometimes unpredictably. The collection of agents and the social context form a complex system where emergent and sometimes surprising behavior arises as each agent strives to adapt and solve its internal constellation of values and needs. It is frequently difficult to trace all the influences that lead to behaviors. As a result, one must make repeated runs (as we show in Part II) to tell if the desired personality or trait tendencies are present. Further, in Part II we train the agents on one data set and test them against a separate dataset to see if these tendencies carry across settings.

2.1) Agent Personality, Emotions, Culture, and Reactions

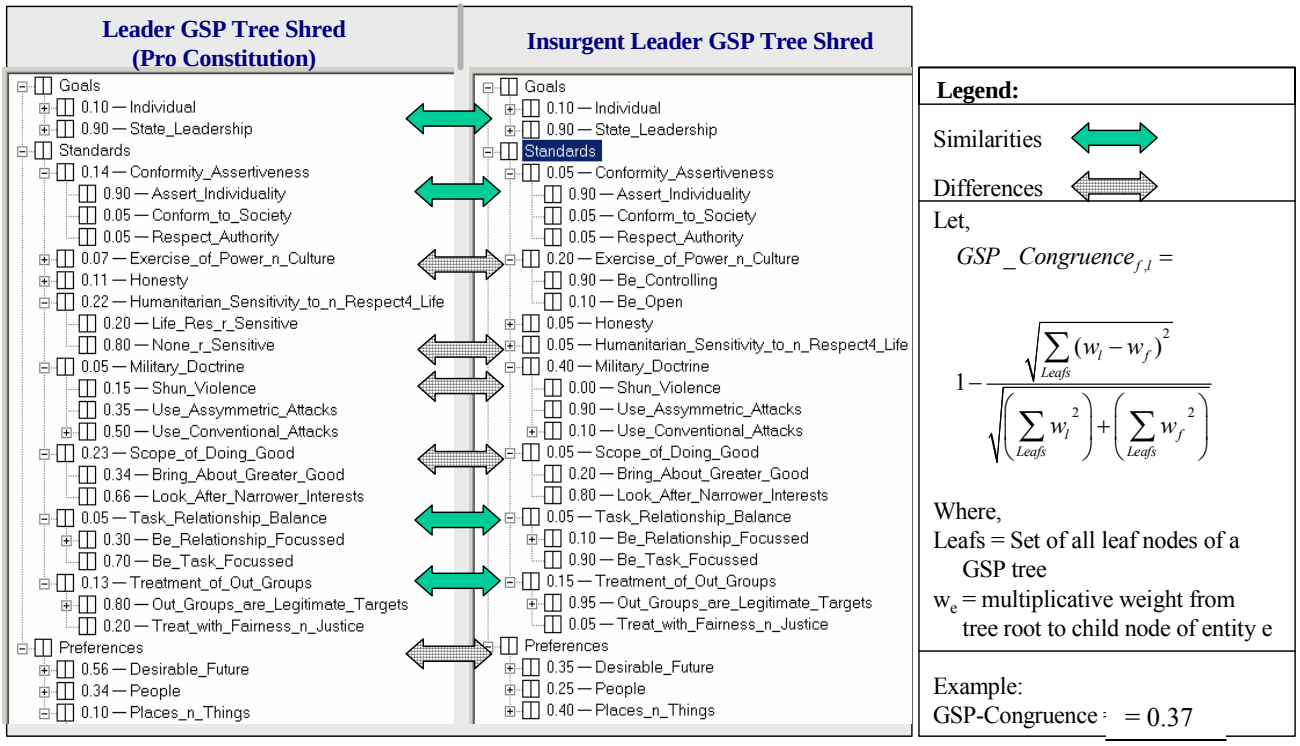
In LeaderSim, each leader is modeled within a framework known as PMFserv (Silverman 2005) where the leader's cultural values and personality traits represented through a Goals, Standards and Preferences (GSP) tree. These are multi-attribute value structures where each tree node is weighted with importance weights. A Preference Tree is one's long term desires for world situations and relations (e.g., no weapons of mass destruction, stop global warming, etc.) that may or may not be achieved in the scope of a scenario. In Lsim agents this translates into a weighted hierarchy of territories and constituencies (e.g., no tokens of leader X in resource Y of territory Z).

The Standards Tree defines the methods a leader is willing to take to attain his/her preferences. Following from the previous section of this article, the Standard tree nodes are mostly Hermann traits governing personal and cultural norms, plus the additions of protocol vs. substance, and top level guidelines related to Economic and Military Doctrine. Also, we add two standards from the GLOBE

study (House, 2004) on scope of doing and sensitivity to life (humanitarianism). Personal, cultural, and social conventions render inappropriate the purely Machiavellian action choices ("One shouldn't destroy a weak ally simply because they are currently useless"). It is within these sets of guidelines where many of the pitfalls associated with shortsighted AI can be sidestepped. Standards (and preferences) allow for the expression of strategic mindsets.

Finally, the Goal Tree covers short-term needs and motivations that implement progress toward preferences. In the Machiavellian and Hermann-profiled world of leaders, the goal tree reduces to a duality of growing vs. protecting the resources in one's constituency. Expressing goals in terms of power and vulnerability provide a high-fidelity means of evaluating the short-term consequences of actions. To this, Athena also adds 3 options for manage reputation (switch from none, to mirroring, to bounded rational) instead of just mirroring in Lsim.

Figure 2 – GSP Tree Structure, Weights and Emotional Activations for Opposing Leaders



With GSP Trees thus structured, we believe it is possible to weight them so that they will reflect the portfolio and strategy choices that a given leader will tend to find attractive. Specifically, the weights adhere to Bayesian mathematics in that all children have a single parent, the weight on all child nodes adds to 1.0, and the weights represent frequency or likelihood in the sense of prior odds. We use differential diagnosis and pairwise comparison to derive these weights, a topic we return to in Section 4 of this write-up. As a precursor to that demonstration and to further illustrate how GSP trees represent the modified Hermann-Hofstede-Globe profiles, consider the left side of Figure 2. There we see the weighted GSP trees for two leaders in Iraq from Spring 2006: the Central Government Leader on the far left and the leader of the insurgency in the middle (Zarqawi). Other papers discuss how the weights may be derived so as to increase credibility: e.g., see Bharathy (2006), Silverman (2002a,b, 2006 pt.2). Here it is more pertinent to discuss how the G-tree implements the Hermann power vs. protect trait. Beneath each subnode that has a + sign, there are further subnodes, but under the G-tree (and P-tree) these are just long sets of constituency resources with importance valuated weights and hence they aren't show here. The standards or S-tree holds most of the traits that constrain choice of actions. It is interesting to see the contrast between the two leaders and where they are more or less similar and different. As expected, the insurgent leader is less about conformity, humanitarianism, scope of doing good, and treating outgroups with fairness, and more about exercise of power, and asymmetric warfare. They are about equal in asserting individuality, honesty, task focus, and lack of emphasis on outgroups. The right side of Figure 2 shows the formula we apply for measuring congruence between two motivational systems. Overall, these two individuals are not that close together, as expected.

According to Ortony, Clore, and Collins, 1988, our emotions are arousals on a set of values (modeled as trees) activated by situational stimuli as well as any internally-recalled stimuli – Appendix I shows how we implement these stimuli and their effects act as releasers of alternative emotional construals and intensity levels, and they assist the agent in recognizing problems, potential decisions, and actions. According to the theory, the activations may variously be thought of as emotions or subjective (moralistic) utility values, the difference being a matter of semantic labeling. Within such a framework, 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.

2.3) Agent Decision Making

What is missing in the previous section is how an agent notices the game world, moves of others, and sense of situation. This discussion will illustrate how this happens using one Hermann factor (power and

vulnerability) as an example. Hopefully, it is fairly straightforward for the reader to extend that to how the other factors are also deployed. Full details exist in Johns (2006).

Central to a given leader's G-Tree reasoning is its perceptions of who threatens it and/or whom it's vulnerable to. Likewise a given leader may be equally interested to estimate who can it influence to best increase its resource assets and thereby its power in the world. Obviously, GSP tree weights will govern how aggressively a given leader pursues each of these vulnerability vs. power concerns, however, we assume that all leader agents need to be able to compute how vulnerable and/or powerful they are at each turn of a game. Since the game rules define precisely which resources can be used to take hostile actions against which other resources, one can derive a measure of a player's *vulnerability* directly from the state of the game world and the rule set. Intuitively, by factoring vulnerability into the world utility calculation, an agent can avoid world configurations in which another is poised to conduct a devastating attack. Adding border defenses, stocking up on supplies, and pulling money out of the economy can all be viewed as behaviors motivated primarily by vulnerability management.

The vulnerability formula (β) works by generating the percentage of a given player's tokens that can be expected to be lost to a given player in the coming round of attack actions (ai). For each hostile action ($a_i \in A$) that can be initiated by another player (g), the number of tokens available to attack and defend is tallied. From this the probability of victory is determined, and then multiplied by the percentage of tokens vulnerable to this attack versus the total number owned by the vulnerable player in each resource category. This is the expected percentage of tokens to be lost if this attack occurs in the next round. The maximum over all attacks, then, gives this player ℓ's vulnerability score β to player y.

$$\beta xy = \text{Max}\, a \in A \left\langle \Pr(a) * \frac{\sigma(x,a)}{C(x)} \right\rangle \qquad [1]$$

Agents who purely manage vulnerability, while interesting in their behavior, are not entirely realistic. Human players tend to balance vulnerability against its inverse, *power*. Where vulnerability measures the expected number of tokens a player can lose to other players in the coming round, power measures the expected number of tokens a player can take from others. The calculation of the power heuristic is exactly the opposite as for vulnerability. Player A's vulnerability to Player B is the same as Player B's power over Player A.

Taking the leader's perceived difference between power and vulnerability provides a surrogate for the leader's overall sense of utility of the current state of the world, G, when divorced from his value system and other factors:

$U_l(Gx) = \alpha x - \beta x$ [2]

Recall, however, that a given leader agent (1) tracks who is aligned with whom, tallying things like trust, (2) monitors all resource levels and who used what actions upon them, and (3) its own actions to achieve its long term preferences or P-tree, as modulated by its standards. Thus α and β serve primarily as activations on the leaf nodes of some of the GSP tree branches. PMFserv uses a wide assortment of similar activation mechanics for other factors and computes the Expected Utility (EU) of the world and of new action possibilities when projecting next steps. That is, PMFserv serves as the point where diverse GSP personality and cultural value sets, stressors, coping style, memories, and perceptions are all integrated into a decision for action (or inaction) to transition to a new state (or remain in the same state) and to determine the portfolio of strategies-moves-actions that best maximize that agent's GSP Tree values as mfollows.

$$\text{Max EU(a)} = \sum_{\text{b}\in\text{Ba}} \text{U}l(\text{b}) * \text{pr(b)} \qquad [3]$$

Where,

a ∈ A

A = action set available after GSP and stress-constrained perception

$p_r(b)$ = probability of action *a* leading to state *b*

$$u_1(b) = \frac{\sum_{k\in K} \xi_k(b)}{11} \qquad [4]$$

$$\xi_k = \xi_k(b \in B) = Instensityofactivation(seeAppendixI)$$

Utilities for next actions, a_k, are derived from the activations on the GSP trees in the usual manner as in Silverman, et al. (2002a, b) and as briefly summarized for power and vulnerability here. That is, utility is the simple summation of all positive and negative activations for an action leading to a state. Since there will be 11 pairs of oppositely valenced activations in PMFserv's emotion model, we normalize the sum as follows so that utility varies between –1 and +1. Interestingly, the 11 pairs of oppositely valenced emotions (and the corresponding activations), correspond to 'swing weights' that capture the impact of the range from best to worst outcomes: e.g., see Watson & Buede (1987).

2.4) Formulating Speech Acts

A large part of the value of our paper-based multiplayer game comes from the interaction among players. Indeed, while the game mechanics provide a concrete scenario with quantified parts, it is the interpretation of board configurations and resulting relationships among players that are of primary

importance. A large part of this interaction concerns the communicative acts and conversations that ensue. Another finding was that we could sort the hundreds of speech acts across several dozen sessions into a rather simple taxonomy consisting of agreements, threats, and statements of motivation. In Athena we have added a human to human chatting window as well as a contract language for sending and receiving speech acts to the agents. Also, espionage budgets can be allocated to try and eavesdrop and learn which agents are conspiring and possibly even the content of some of their messages.

An Lsim or Athena agent can invoke PMFserv to compute the expected utility of agreement, threat and/or motivation statements. For example, for threats, supposing ℓ_2 can perform an action of higher utility to ℓ_1 than could be done otherwise, then it is worthwhile to search for actions that can induce ℓ_2 to perform it. The ℓ_2 then searches for actions that ℓ_1 can perform, minimizing the utility to ℓ_2. This allows it to calculate the utilities of heeding and ignoring the threat for both parties.

$$U_{heed} = EU(a_0(\ell_1), a_1) \quad [5]$$

$$U_{ignore} = EU(a_2, a_0(\ell_2)) \quad [6]$$

where $a_0(\ell_1)$ and $a_0(\ell_2)$ are the chosen actions for ℓ_1 and ℓ_2, respectively. If the utility of heeding the threat is greater than ignoring it for both parties, the threat is worth carrying out.

2.5) Intention Modeling and Mirroring Biases

In order to assess speech acts and motivations of others, agents must model and learn about other players in the game. For computational efficiency sake, the agent-unique GSP trees and universal GSP activations from world affordances are convolved (along with that agent's starting relationship matrix and various group properties) before run time into a vector that if we were interested in evolution would be called the DNA of each agent (i) -- a set of equality constraints or vectors of probabilistic tendencies for all combinations of actions on all possible world objects DNA{[G-gene][S-gene][P-gene]}. Since we focus on real world personalities, we are more interested in current agent learning and adapting than in evolution. An agent's learning occurs when he forms a model of another's "DNA" (or when he updates his own). According to the Hermann trait set, this can vary as a function of an agent's Conceptual Complexity, a trait we allow to assume one of three levels, low, medium, and high as follows – see Silverman et al (2007a) for more detail:

- Worst Case – An agent is unaware of the DNA of others, and instead models them by viewing their alignments (relationship matrix), action possibilities, and resource levels. The agent then computes Vulnerability and Power [1, 2] to assess the worst that others can do to him and/or the worst he can do to others, choosing whatever maximizes his expected utility. Depending on other agents' hiding resources and on the agent's stress level, he may not always have enough information and/or time to do these computations thoroughly or accurately.
- Mirrored – An agent forms an initial model of each other leader's DNA by mirroring his own DNA through the alignment matrix (and current trust vector). Thus, if agent X wants to eliminate agent Y from a certain part of the world, and Y is an enemy, then X mirrors that Y must want to eliminate him. This is then used to modify Power and

Vulnerability in finding the best response. A mirroring agent only updates DNA models with information that confirms them: see Johns (2006).

- Informed –An agent models his adversaries by parsing the DNA into a linear system of inequalities with known terms but unknown weights. By introducing slack variables to impose non-negativity conditions, these equations become equalities and in matrix notation, this will be **Hy = 0**, where the matrix **H** is a coefficient matrix and **y** denotes the vector of unknowns. A reasonable first approximation to the y unknowns may be derived via simplex method if at least 3 opponent decisions are observed. Further observations narrow the region of uncertainty surrounding these estimates of the opponent's value system. This method is intensive and is only applied to model an agent's worst opponent. Also, we do not believe this is how humans do this task, and are researching more descriptive as well as efficient solutions: see Bharathy (2006)

3) Modeling Follower Value Systems: The Evolution of Dangerous Ideas

We introduce three refinements in order to also be able to model the values and motivations of followers – (1) additions to the GSP trees, (2) a group-affinity profiling instrument, and (3) group transfer dynamics. In keeping with Aim 2, each of these refinements is an implementation of a well-respected model drawn from the social sciences. In this section, we describe this integration and synthesis in terms of an example individual Villager from a SE Asian country. He grew up as a Muslim in a society with a Buddhist majority (Blueland on right of Figure 3), where the Buddhists occupy the cities, hold the elected posts, and get the best jobs. Most of the Muslim's are villagers in the rural provinces, a region they would like to separate and have autonomy over (shown as Yellowland), The elder villagers tend to be more moderate and compliant, while the younger ones, like the one modeled in Figure 3, were sent to Wahhabi schools elsewhere, went to college for training, and have returned to find there are no sources of employment that will use their college training (Yellow's economy is very low in Figure 3). Further, their religious practices are frowned upon, their community is unable to open Wahhabi schools officially (these are being operated by the returned college grads in their homes), and their initially peaceful protest events have been met with police brutality (Blue ArmedForces present in Yellowland). The bottom left box of Figure 3 shows the country leader has been discriminating against Muslim villagers, the moderates disagree with these policies, while this individual is opting to "oppose". He is just at the juncture of having switched groups from moderate villagers to the local insurgency (Red tokens on right side in YellowLand) which itself is on the verge of connecting with groups such as Jemaah Islamiah and Al Qaeda if the discrimination continues. Let us now see how the three sets of refinements allow the PMFserv model to capture such a radicalization.

In terms of the GSP tree changes, as the right side of Figure 3 reveals, there are additional nodes on the Goal and Standards trees. While Figure 3 does not show the full trees, the previous goals and standards are still there but may be zeroed out if this is strictly a follower, or may be left in at some

degree of importance if this is a mid-level leader. For villagers in general, where day-to-day existence is a struggle, Maslow's hierarchy of needs is considered a useful representation of the range of short term goals that a person might have to be concerned about. Each of these nodes are activated by lower level branches on the tree, tanks in the physiology and stress model, or the affinity instrument described below. Clearly all of this individual's basic needs are failing to be met, and the Malovian nodes are all activated with negative stimuli in this example, accounting for the extreme distress level of the goal emotions on the left of Figure 3. In terms of the Standards Tree, the Hermann factors are all still there, but two new nodes on the tree exist that capture several Hofstede (2003) and House (2004)'s GLOBE study cultural factors. Conformity Assertiveness is a way to capture Hofstede's Power-Distance and Individualism factors (respect authority, conform to society) with the GLOBE study's Assertiveness factor. Likewise, the Humanitarian node is a GLOBE factor. The weights on Figure 3 for this villager's standards reflect his idealism, training in religious and college settings, and his alienation. Hence he differs from the moderate villagers in the various Hofstede and GLOBE factors, and also his personal 'military doctrine' does not shun violence. According to Sageman (2004), these are the seeds, but they do not alone explain performing violent and suicidal acts. The activations on his Standards tree tanks must also be such to tip him from just disagreement and protest over to actual physical opposition and violence. We turn now to the second set of refinements that enable radicalization and provide the needed extra activations for these tanks.

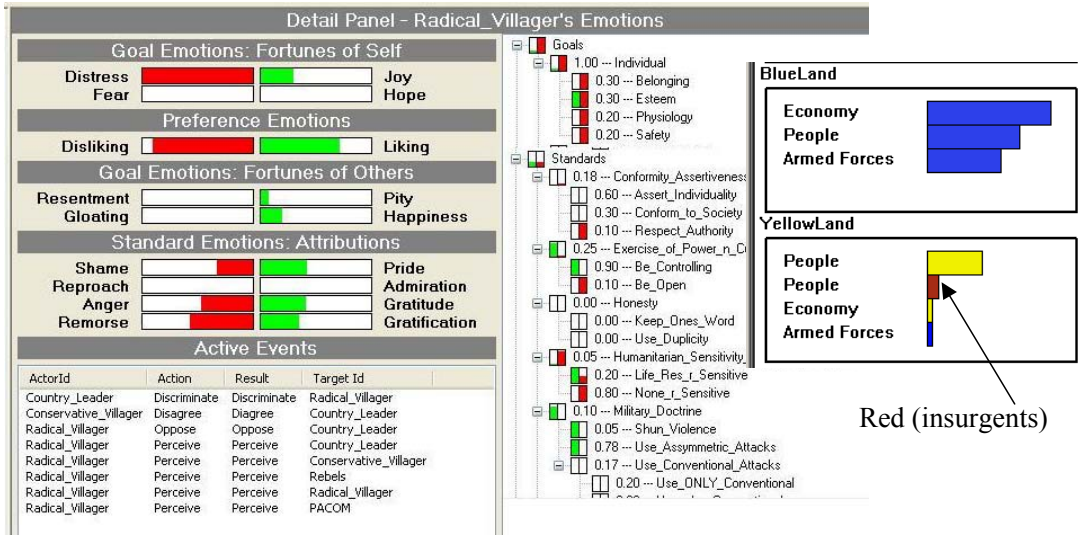


Figure 3 – Screens from Lsim and PMFserv Depicting a Buddhist-Muslim Conflict in SE Asia

Specifically, for determining an individual's group affinity, one needs an instrument that measures it. The instrument that we have adopted here involves Eidelson and Eidelson (2003) who have developed a five belief ("dangerous ideas") framework for better understanding the psychology of

individual-group dynamics particularly relevant to conflict settings. These five beliefs revolve around issues of vulnerability, injustice, distrust, superiority, and helplessness(in short H|VID|S):

•*Vulnerability* (V) Revolves around a sense of living in harm's way amid constant threat and peril.

•*Injustice (I)*. perception of being victim of mistreatment by specific others or by the world at large.

•*Distrust (D)*. the presumed hostility and malicious intent of other individuals or other groups.

•*Superiority (S)*. conviction of being better than others—morally superior, chosen, entitled.

•*Helplessness (H)*. Refers to perceived inability to influence or control events and outcomes; self-perpetuating because it diminishes motivation.

Eidelson has developed the Individual-Group Belief Inventory (IGBI) to measure these five beliefs. These beliefs are considered particularly important influences on a group member's perceptions of his/her group's current circumstances and future prospects. In this writeup, we will only focus on how we combine the vulnerability, injustice, and distrust domains into one broader domain (hereafter labeled VID) representing a member's belief that his/her group has legitimate grievances against a threatening adversary. Thus, $\forall$ (F,L), (L,L):

Vulnerability: $V = (e_{distress} + e_{dislike}) - (e_{joy} + e_{liking})$

Injustice: $I = e_{anger} - e_{gratitude}$

Distrust: $D = c_1 e_{reproach} - c_2 e_{admire}$

VID: $VID = V + I + D$

Where,

c_1, c_2 – Weight constants for prime (c_1) and secondary (c_2) emotions, currently (1, 0.1) respectively

e_x = emotional activation from PMFserv

Depending on the perceiver, two individuals may view the same group as superior or inferior, as suffering grave injustices or as exaggerating minor slights, as helpless or capable of effective action, and so on. The markups expose all the possibilities and include universal rules that apply to all viewers. Thus if a viewer sees a group as vulnerable and he doesn't want to be vulnerable, there will be negative activations for remain loyal to this group and following its action choice policies. Conversely less negative (and possibly positive) activations will be afforded for the member who exits away from this group perceived as vulnerable. The mechanics of how PMFserv handles perception have been covered in Cornwell et al. (2003), Silverman et al. (2005, 2006a) and we won't go into them in full detail here except to mention that once markup is done, perceptual types are driven by the agent's emotional and utility construals of the current state of the world. For example, when Standards and Preferences are perceived to be violated Injustice and the anger emotion are aroused, while Vulnerability comes from the Safety

failings on the GSP tree, and emotions about that. Part of a Muslim villager's perception of the Bhuddist country leader is as shown in the PMFserv screen of Figure 4 which reflects a range of VID parameter settings (neutral, very low, all the way to High), and the fact that the villager sees the leader as causing VID High. This perception on the part of the villager is based on observing actions that the leader has taken to discriminate against Muslims (no jobs), to deny their request for Wahhabi schools, to brutally put down protest, etc. Conversely, Buddhists who view this leader see him as neutral, or low VID. The former would be for those in his political party, while the latter would be others who pursue normal political processes to resist his positions. Likewise, the Muslim villagers and Buddhists will form similar judgments about the various Muslim groups/leaders in the scenario.

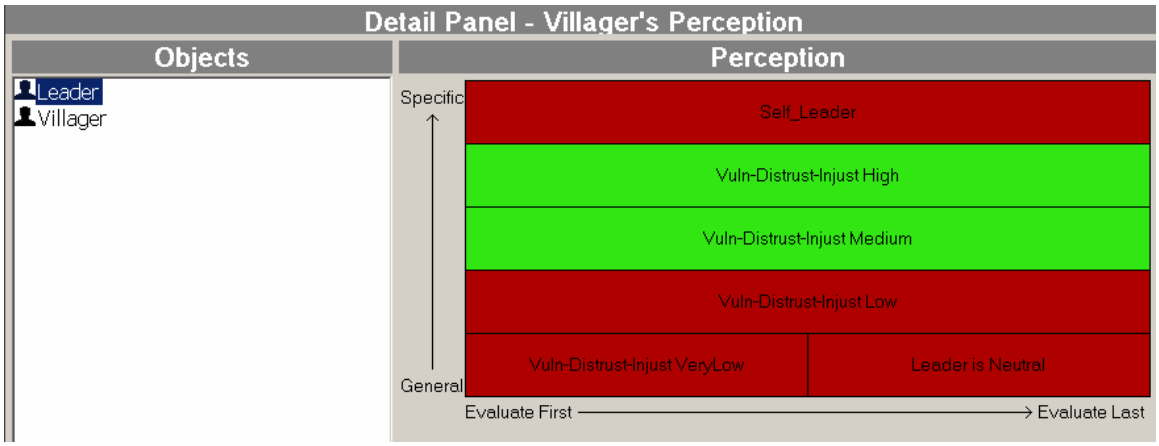


Figure 4 – Example of a Villager's DI Perception Possibilities When Viewing a Given Leader

Up to this point we have seen how an agent's GSPs and perception of the range of possible DI markups permit him to examine the state of the world and assess the utility and emotional construals that each group/leader affords up to that point. What remains is to examine the third piece, the parameters that govern switching group affinities, and for that we implement a Hirshman (1970) type model on exit, voice, and loyalty. Graphically, one may imagine a vertical axis with VID increasing and the horizontal axis as members of that group. The loyalty for remaining in that group is governed by the slope or elasticity of the downward sloping demand curve. As VID rises, members may express their grievance (GR) as voice, if their grievance tolerance passes a threshold, they may chose to exit the group. Hirshman's research concerned consumers who found exit less costly than voice (complaints). In a political world, exit is more costly, and hence voice to complain and improve situations will be of import.

Mathematically, the reader may recall ϕ (r_{ij}) from earlier Equation [1]. Here we examine the case where j is a group (or leader) and the term refers to the membership, relationship, or strength of affinity of agent i to group j. An agent i can belong to multiple groups at varying strength according to,

$$\Phi(r_{iA}) = Superiority_A \times GSPcongruence_{iA} / VID_{Ai} \quad [7]$$

where Superiority and VID are from DI instruments if available, else derived by GSP trees of agent I in reacting to leader or group A. Groups are characterized by GSP weights for the average of all members as well as by property lists defined a priori (religion, political system, etc.), and GroupPorosity factors. GSP congruence is estimated using the sum of the means square differences in the GSP nodes. GSPcongruence = 1 – Sqrt[Sum[(wi1 –wi2)^2], which is the correlation of the weights between two GSP trees. If an agent is in Group B, it will not be drawn to a Group C whose GSP archetype is substantially incongruent to its own. If an agent is in a group (or under control of a leader) whose average GSP is greatly different from its own, the agents tend to use Voice to resist the leader or attempt to Exit to another group, depending on porosity. Let us now see how each of these work.

If agent i desires to exit from any group A to join any C, this is governed by the delta in utility of membership in each group plus a cost factor adjusted for transfer rate or demand elasticity. If the delta is positive, or larger than some loyalty factor, exit may occur. Let, this delta be

$$\Delta\Phi_i = [U(\Phi_C) + COST_{TR} / TR_{AC}] - U(\Phi_A) \quad [8]$$

where,

$U(\phi)$ = utility of membership, found by invoking the PMFserv emotion model and GSP trees

$COST_{TR}$ = cost of migration, land costs, and lost opportunity costs

TR_{ij} = Transfer Rate or group porosity, a measure of ease of entry to or exit from group j for agent i. (TR is a negative value that grows larger as porosity grows)

$$TR_{A \to C} = Salience_{ExitA} \times Salience_{EnterC} \times GSPcongruence_{iC} \quad [9]$$

This is the transfer rate and it varies between (0,-1). Salience is the extent to which a group permits exiting by ingroup members, and entry by outgroup members. It is the porosity permitted by the group. There is a tuple or value pair that gives both salienceForEntry and salienceForExit. The demand elasticity for exiting a group is 1/TR.

Assuming COST is too high or TR is too small and an agent cannot exit, then their options are to remain silent or use Voice to express their grievance over VID levels. Let, GR_{ij} be the expressed grievance level by agent i relative to group j. This is computed as the negative emotions activated when agent i perceives the VID levels of a given group or leader's actions. GR is thus feelings about the world due to actions that violate one's GSPs and that cause negative emotional utility. In the current model, expressed Grievance is fed from the levels of the goods and security tanks and the individual actions of

the leader to affect what these contain. Of course, perceived Grievance alone does not make a rebellion, but it might contribute towards the rebellion. The Grievance should be combined with a feeling that the other group or in this case, the leader or the central authority, induces helplessness, vulnerability, injustice, distrust, and/ or attack on superiority. As the leaders did with Figure 1, the followers similarly take each set of opposing groups and place them along a scale as shown below. The decision that the villagers make is expressed as grievance, where the grievance is in the scale of –4 to +4 are given below (also shown are the Grievance State IDs of the simulation of Sect 4.2):

←				Villager Decision				→
Sacrifice, Go on Attacks	Support, Vote for GroupA	Join Authority Group A	Agree	Neutral (undecideds in Group B)	Disagree, Vote against A	Join, Opposition Group C	Oppose, Non-Violent	Fight Rebel, Exit A
-4.0	-3.0	-2.0	-1.0	0.0	+1.0	+2.0	+3.0	+4.0
				GS0	GS1	GS2	GS3	GS4

These actions are on abstract scale, which ranges from total support of the majority that is oppressing you (if you can't lick them, join them), to being undecided and/or helpless in the middle, to the other extreme of supporting and ultimately exiting A and joining the insurgency. At the extremes on either end, the agent will submit to militaristic commands of the leader of that group, while at the next level two lower levels they will be only willing to go to protests, and verbally and economically support the activities of that group's leaders. Thus, every state all the way through GS4 represent Voice. We only permit Exit from A and joining of C after occupying GS4 for a significant interval.

In terms of our Villager from earlier Figure 3, he has just transitioned from the undecided or neutral Muslim villagers category to the OPPOSE level (GS3). He saw Country Leader as creating high VID. The insurgency group, however, has high GSPcongruence for him and low salienceForEntry given his religious training. This transfer of primary group membership would afford him relief from his sense of helplessness, and we see his military nodes on his GSP tree getting activations as a result. Since he is not yet at the level of performing missions, both his shun violence and asymmetric attacks are simultaneously activated.

4) Existence Proof: Implementing the Leader-Follower Game and Agent

Section 1 presented a theory of role playing socio-cultural games that guides various implementations we have attempted to date. These implemented games themselves serve as an 'existence proof' that the syntheses discussed here can be successfully integrated and deployed. The socio-cultural game theory summarized in Section 1 and the leader-follower agent framework of the rest of this article provides a

point of departure for delineating the 5Ps of any scenario – i.e., people (roles, relationships), plot, place, processes (campaigns, actions, speech acts), and player pedagogy. In Part II we review two of these games and how their agents were trained and validated. Here we simply summarize the games authored to date.

To date, we have authored two implementations of this theory that run inside of 3rd party videogame engines:

- Black Hawk Down – Using the Unreal Engine, this includes a recreation of the Bakara Market. Your mission as the player is to lead a squad of rangers (3 other bots) to Mike Durant's crashed helicopter, to chase away looters without upsetting them, destroy the crashed chopper, and return safely to your Humvee. All the crowds that one encounters are PMFserv agents organized in groups of leaders and followers. This effort was done with just the basic PMFserv emotion and relationship model. It is the effort that made us realize we needed the synthesis of this paper. Videoclips of game action exist at www.seas.upenn.edu/~barryg/HBMR.html and articles on it are (vanLent et al. 2004, Silverman et al. 2006b).
- SCALE-UP -- This is an Iraqi street scene where you play the squad leader charged with dispersing the crowd. It is a dialog and cultural sensitivity game. If you don't find out the crowd's grievances (need for food), or if you disrespect their leader because you don't understand simple mores and norms in their culture, then the crowd grows increasingly agitated. Having a coded version of the full follower framework as delineated in this article permitted us to rapidly embed PMFserv behind the BigWorld Engine to drive the crowd behaviors of this game pilot (MacDonald et. al, 2006).

Such 3-D animated worlds are useful for training of street scenes, however, the graphics become an unnecessary burden when trying to analyze and understand what lies behind such conflicts. For more strategic policy analysis and EBO studies, we have authored three implementations of the diplomatic game framework of Section 1.1 as follows. Each of these is itself a game generator and can be used to mock up and study conflicts from around the world. They are:

- LeaderSim (or Lsim) – This is a software prototype that keeps scenarios to 3 territories, 3 resources in each, and only the 10 or so actions of the 1st column of Figure 1. For the leader vs. leader game, this is the smallest implementation of Figure 1 possible so that we can easily build and test all of the core ideas of the theory that pertain to leaders; Silverman et al.(2007) describes it more fully and conducts a validation study using leaders of the Third Crusade.
- Athena's Prism –This is a distributed architecture, multi-player RPG version that has been delivered as a fully functioning computer game (user-defined territories, each with 12 resource categories and 100 actions) to an agency of the US Government in mid 2005, and which has been used to help analysts generate ideas for influencing conflict situations. The operations concept, is that this could be played by

one human on one PC with all the others simulated by artificial intelligence (AI), or all leaders could be human players at remote stations (no AI), or any combination in between of humans and AI, including all AI. The current implementation of Athena also includes a rapid scenario and leader editor module. It is not our purpose here, and there is insufficient space, to review these many screens and features, and the interested reader is referred to Silverman, Rees, et al. (2005) for those details.

- FactionSim - The previous two RPGs have roles for leaders only. The followers are all "poker chips" that can only comply with the leaders' wishes. FactionSim expands the Lsim game to allow hierarchies of PMFserv agents to play rival leaders and followers within a faction (and it allows many factions). It thus is the basic set of all games LL, LF, FL, FF, etc. We have also opened up its three resources (economy, security, and political support) so that outside models can be plugged in and run to make them more realistic (and also to test those models). Part II of this article presents FactionSim and two case study implementations it was used for under DARPA sponsorship (SE Asia and the MidEast).

5) Lessons Learned and Next Steps

This research set out to try and synthesize best of breed human behavior models useful for EBO studies and analyses. This article started with a theory of leaders and followers in factional conflict games, then presented a synthesis of best of breed scientific models for implementing synthetic versions of those leaders and followers including simple runs showing plausible outputs of diverse leader and follower agents' reasoning mechanisms. Discussion finally ended up with an "existence proof" that these agents and indeed the leader-follower theory have guided development of games intended to help trainees and analysts explore alternative courses of actions.

We can conclude that Aim 1 was only partially achieved -- to create a role-playing game generator where one could rapidly set up and play out or study how to influence numerous conflict scenarios from around the world. Indeed, games have been created and fielded that follow the theory presented here. And these in turn, refined the theory, so that our specification for a game generator now reduces conflicts to the bare essentials while preserving realism and allowing one to explore the intertwined issues affecting welfare (economy, in-group standards, health services), security (freedoms/liberties, military), and political support for leaders (popularity of positions). All the game-play we have created and studied gave us a rich source of data to help guide the construction of agents who can serve as synthetic opponents, allies, followers, and the like. Also, we have learned that our game state representations are intuitive and that domain experts can readily use them to express conflict scenarios that are hard to verbalize. As with anything done in software, there are always next levels of sophistication and detail that one can add, and we identified many new features we would like to add such as, to mention a few examples, (2) scale up of all features shown here for the larger game generator we

call Athena's Prism; (2) resources and assets (e.g., economy, media, and black markets) that are supported by third party models of their dynamics; and (3) a "dashboard" that helps users to design what-if experiments, generate reports on model outcomes, and explains agent decision choices and effects. These are some of the next steps for the game generator.

Also many larger challenges remain. For one thing, the generator, is not just one piece of code, but is a guideline for factional games, a toolset (PMFserv) for deploying leader and follower agents into those game worlds, and various implementations attempted to date. More significantly, the goal of an EBO game generator that is truly realistic and valid is a long term challenge problem for the scientific, software, and EBO communities that may be years away. The research on generators such as Lsim, Athena, FactionSim serve to illustrate how hard this is, and how any one generator really only tackles a slice of the problem. In Part II of this article we revisit this issue and further examine the need for open framework that can serve as a plugin testbed of many peoples' contributions – resource models, DIME-PMESII operations simulators, and agent models. FactionSim will be presented there as a way for our leader and follower syntheses and software to bridge to the many other partial solutions that others in the field have to offer.

As to Aim 2 of this paper, modeling leaders and followers is a complex enterprise and one would like to use only first principles of social science, yet that field also has not matured sufficiently. Still, that is no excuse for modelers to "make up" their own rules and algorithm for how groups behave, nor is it justification to just create entertaining agents. The alterative we explored here is to try and adopt best-of-breed and well-respected social science models for leadership, group dynamics, and the hearts and minds of the populace. These models are implemented atop a unified architecture of cognition, call PMFserv that manages six modules of an agent's mind: memory, perception, physiology/stress/coping level, value system, and emotional construal, relationships and models of other, and (stress and emotion-constrained) decision making processes. PMFServ exposes many parameters in each of these modules and permits analysts/developers to visually "program" best-of-breed social science models that govern how the modules work, and in turn, how that agent tends to behave. This framework supported the ready implementation of leader models from Hermann (style), Hofstede and Globe (cultural factors), and Heuer (biases) atop pre-existing models in the PMFServ modules. Likewise, the PMFServ modules allowed group followers to be readily modeled via their personal motivations (Maslow-style), their perception of factions and leaders (Eidelson Dangerous Ideas Model), and action choices ranging across loyalty, exit, and voice (Hirshman Model).

We thus enabled a synthesis and by that exposed what is missing (Aim 3). For both the leader and follower models, we showed screens with simple tests that indicate the agents seem to be using these models appropriately. Readers could see how synthesized versions of Iraqi leaders' GSP tree values and

emotions drove them to make action choices in their own ingroup's interests – interests that lead to a conflict over control of resources in that nation. For a different conflict, readers also could see how a moderate Muslim follower was radicalized and shifted toward the insurgency. In short, inside the PMFserv agents, one can readily observe and track their GSP tree implementation of Maslow, Hermann, Hofstede, and other factors, and preference functions. One can follow how they update the Eidelson model factors of group and leader achievement, and how they reach Hirshman-type decisions concerning group membership and actions to take. This convinces one that the synthesis of complementary scientific models "works" at a mechanical level, and appears to make sense. But without larger assessments and out-of-sample testing of the synthesis are we just guilty of over-fitting with all these models? Part II of this article turns to the questions of whether these syntheses collectively pass tests of validity? If so, they minimize the dimensionality curse and reduce the internal parameters considered as independent variables for EBO studies and analyses. Thus the EBO studies can proceed unencumbered.

REFERENCES

Armstrong, JS (2002). Assessing game theory, role playing and unaided judgment. *International Journal of Forecasting*, 18, 345-352.

Bharathy, G (2005) *A knowledge engineering based systems methodology for integrating of social science frameworks for modeling agents with cognition, personality & culture*. (dissertn, U. Pennsylvania, 2006)

Buckley, W. (1967). Sociology and modern systems theory. Englewood Cliffs, N.J.: Prentice Hall, Inc.

Camerer, C (2003), *Behavioral Game Theory*, Princeton: Princeton Univ. Press.

Canella, A. A., & Monroe, M. J. (1997). Contrasting perspectives on strategic leaders: Toward a more realistic view of top managers. Journal of Management, 23(3), 213-230.

Chemers, M. M. (1997). An Integrative Theory of Leadership. Mahwah, NJ: Lawrence Erlbaum.

de Marchi, S, (2005), *Computational and Mathematical Modeling in the Social Sciences*, Cambridge Univ. Press.

Cilliers, P. (1998). Complexity and postmodernism: Understanding complex systems. London: Routledge.

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

Geib, C. (2004). Assessing the complexity of plan recognition. Proceedings of AAAI-2004.

Green, K. C. (2002). Forecasting decisions in conflict situations: a comparison of game theory, role playing and unaided judgment. *International Journal of Forecasting*, 18, 321–344

Hermann, M. G. (1999). Assessing leadership style. Hilliard, OH: Social Science Automation, Inc.

Heuer, R. J., Jr. (1999). *Psychology of Intelligence Analysis*. Washington, DC: Center for the Study of Intelligence, Central Intelligence Agency.

Hirschman, A.O. (1970). *Exit, voice, and loyalty.* Cambridge, MA: Harvard University Press.

Hofstede, G., (2003), *Culture's Consequences, Comparing Values, Behaviors, Institutions, and Organizations Across Nations* Newbury Park, CA: Sage Publications. www.geert-hofstede.com

House, RJ, Hanges, PJ, Javidan M, et al., (2004), Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies, Thousand Oaks, CA: Sage Publications

Johns, M., *Deception and Trust in Complex Semi-Competitive Environments* (dissertn, U. .Penn, 2006)

Kaminka, G. A., & Avrahami, D. (2004). Symbolic behavior-recognition. 3rd International Joint Conference on Autonomous Agents & Multi Agent Systems (AAMAS). NYC, July 19-23, 2004.

Kautz, HA, Allen, JF (1986), *Generalized Plan Recognition*, in AAAI 86, Mtn View: AAAI, pp. 32-37.

Kojima, H. (2004, June), Interview, *Game Informer*, p.42.

McDonald, D, Lazarus, R., Leung, A, et al. *Interoperable Human Behavior Models for Simulations,* 15th Conf. On Behavioral Representation in Modeling & Simulation (BRIMS), SISO, May. 2006

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, DC: National Academy Press.

Silverman,B.G. (2005). Human performance simulation. In J.Ness, D.Ritzer, & V.Tepe (Eds.), *The Science and Simulation of Human Performance (Chapter 9)*. New York: Elsevier.

Silverman, BG, Rees, R, et al. (2005, May), "Athena's Prism: A Diplomatic Strategy Role Playing Game for Generating Ideas and Exploring Alternatives," in Proc. Internat'l Conf on Intelligence Analysis, MacLean, VA: Mitre.

Silverman, B. G., Johns, M., Cornwell, J., & O'Brien, K. (2006a). Human Behavior Models for Agents in Simulators and Games: Part I – Enabling Science with PMFserv. *Presence* , v. 15: 2, pp. 139-162, April.

Silverman, B. G., O'Brien, K., Cornwell, J. (2006b). Human Behavior Models for Agents in Simulators and Games: Part II – Gamebot Engineering with PMFserv. *Presence,* v. 15: 2, pp. 163-185, April.

Silverman, BG, Bharathy, G, Nye, B, et al., (2007) "Socio-Cultural Games for Training and Analysis", IEEE SMC, pending publ.

Simari, G. I., & Parsons, S. (2004). On approximating the best decision for an autonomous agent. 3rd AAMAS Conf., W16: Workshop on Game Theory and Decision Theory. NYC.

Swedberg, R., *Sociology and Game Theory,* J. Theory and Sociology, 30:3, 301-335.

Uhl-Bien, M., Marion, R., & McKelvey, B. (in press). Complexity leadership theory: Shifting leadership from the industrial age to the knowledge era. *The Leadership Quarterly*.

van Lent, M., McAlinden, R., Brobst, P, et al. (2004). Enhancing the behaviorial fidelity of synthetic entities with human behavior models. *13th Conf Behavior Representation in Modeling & Simulation (BRIMS),* SISOstandards.org, Arlington, Virginia, May 12-15, 2004.
Watson, S., Buede, D, (1987), *Decision Synthesis*, Cambridge: Cambidge University Press
Woodcock, S. (2002). Game AI: The state of the industry 2001-2002. *Game Developer, 9*(7), 31.

ACKNOWLEDGEMENT: This research was partially supported by the US Government (paper game, game engine, leader models), DARPA (follower models, case studies), AFOSR (FactionSim), and by the Beck Scholarship Fund. Also, we thank the US Government for their guidance and encouragement, though no one except the authors is responsible for any statements or errors in this manuscript.

APPENDIX I

Just to the left of the weight value on each node of the GSP trees of Figure 3 are two "reservoirs" that reflect the current activation of success and failure of this node, respectively. These reservoirs are activated and filled by events and states of the game world as observed by the agent. Figure 3 shows the start of a game where Saladin has yet to be attacked and he perceives the world in a fairly satisfied light. In general, we propose that any of a number of *k* diverse activations could arise with intensity, ξ, and that this intensity would be somehow correlated to importance of one's GSP values or node set (GSP) and whether those concerns succeed or fail for the state in question. We express this as

$$\xi_k(b \in B) = \sum_{j \in J_k} \sum_{v \in V} \left[Wij(v \in V) * \Phi(r_j) * \zeta(v) * \psi \right] \qquad [1]$$

Where,

$\xi_k \rightarrow \xi_k(b \in B)$ = Intensity of activation, *k*, due to the bth state of the world.

J_k = The set of all agents and objects relevant to *k*. J_1 is the set consisting only of the self, and J_2 is the set consisting of everyone but the self, and J is the union of J_1 and J_2.

$W(v \in V)$ = Weighted importance of value set V to the agent.

V = The set of goals, standards, and preferences held by the agent.

$\phi(r_j)$ = A function that captures the strength of positive and negative relationships one has with agent or object j that are effected or spared in state b.

$\zeta(v)$ = degree of activation for a goal, standard, or preference

ψ = A function that captures temporal factors of the state and how to discount (decay) and merge one's GSP activations from the past (history vector), in the present, and for the future

It is important to note that the weights adhere to principles of probability; e.g., all child node insights add to unity beneath a given parent, activations and weights are multiplied up a branch, and no child has multiple parents (independence). Although we use fixed weights on the GSP trees, the reservoirs serve to render them dynamic and adaptive to the agent's current needs. Thus, when a given success reservoir is

filled, that tends to nullify the importance of the weight on that node (or amplify it if the failure reservoir is filled). In this fashion, one can think of a form of spreading activation (and deactivation) across the GSP structure as the scenario proceeds.