University of Pennsylvania
ScholarlyCommons

Center for Human Modeling and Simulation | Department of Computer & Information Science

5-1-2001

How Emotions and Personality Effect the Utility of Alternative Decisions: A Terrorist Target Selection Case Study

Michael Johns
University of Pennsylvania

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

Postprint version. Published in *Tenth Conference On Computer Generated Forces and Behavioral Representation*, May 2001, 11 pages.
Publisher URL: http://www.sisostds.org/index.php

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/hms/10
For more information, please contact repository@pobox.upenn.edu.

How Emotions and Personality Effect the Utility of Alternative Decisions: A Terrorist Target Selection Case Study

Abstract
The role of emotion modeling in the development of computerized agents has long been unclear. This is partially due to instability in the philosophical issues of the problem as psychologists struggle to build models for their own purposes, and partially due to the often-wide gap between these theories and that which can be implemented by an agent author. This paper describes an effort to use emotion models in part as a deep model of utility for use in decision theoretic agents. This allows for the creation of simulated forces capable of balancing a great deal of competing goals, and in doing so they behave, for better or for worse, in a more realistic manner.

Keywords
emotion, decision theory

Comments
Postprint version. Published in *Tenth Conference On Computer Generated Forces and Behavioral Representation*, May 2001, 11 pages.
Publisher URL: http://www.sisostds.org/index.php

This conference paper is available at ScholarlyCommons: http://repository.upenn.edu/hms/10

How Emotions and Personality Effect the Utility of Alternative Decisions: A Terrorist Target Selection Case Study

Michael Johns[2]
Barry G. Silverman, PhD [1,2]
1- Systems Engineering Dept. & Inst. For Research in Cognitive Science (IRCS)
2- Center for Human Modeling & Simulation, Computer & Info Science Dept.
University of Pennsylvania,
Philadelphia, PA 19104-6315
215-573-8368
mjohns@seas.upenn.edu
barryg@seas.upenn.edu

Keywords:
Emotion, Decision theory

ABSTRACT: *The role of emotion modeling in the development of computerized agents has long been unclear. This is partially due to instability in the philosophical issues of the problem as psychologists struggle to build models for their own purposes, and partially due to the often-wide gap between these theories and that which can be implemented by an agent author. This paper describes an effort to use emotion models in part as a deep model of utility for use in decision theoretic agents. This allows for the creation of simulated forces capable of balancing a great deal of competing goals, and in doing so they behave, for better or for worse, in a more realistic manner.*

1. Introduction

Our broad research goal is concerned with demonstrating how existing human behavior modeling frameworks can be effectively synthesized and deployed in agent decision processing [1]. A particularly important goal is to illustrate how these models help systematically capturing and portraying individual differences in socially intelligent agents. For example, how can agents be created to systematically reflect contextually relevant emotions and personality, and further, how do these affect their decision making behavior.

Within the military training domain, these research issues manifest themselves as the desire to be able to dial up different opponent groups against which to train, e.g. the Iraqi Republican Guard, the Hamas-style suicide bomber, or the clandestine minions of Bin Laden.

The idea that humans are rational actors whose decisions are often clouded by emotion is as old as Western thought. Until recently, artificial intelligence research concentrated primarily on the "rational" aspect of this, reasoning that since the problem of making good decisions is so difficult in itself that to introduce emotion into the equation would make the performance of the agent even worse. Recent theories e.g. [1, 2, 5], however, suggest a quite different relationship: that emotions are a vital part of the decision-making process that manage the influence of a great deal of competing motivations. According to these theories, integrating emotion models into our agents will yield not only better decision-makers, but also more realistic behavior by providing a deep model of utility. These agents will delicately balance, for example, threat elimination versus self-preservation, in much the same way it is believed that people do.

To begin to model this computationally, it is first necessary to model how emotions come about. A variety of tools are available from the psychology literature, particularly a class of models known as cognitive appraisal theories. These include the models of Lazarus [5], Roseman [9], and Ortony, Clore, and Collins (OCC) [7], and take as input a set of things that the agent is concerned about and how they were effected recently, and determine which emotions result. For example, the OCC model separates concerns into goals (desired states of the world), standards (ideas about how people should act) and preferences (likes and dislikes). These are evaluated against the current state of the world, and some mixture of twenty-two emotions results. A key idea of these theories is that emotions are intrinsically valenced – they can be identified as being desirable or undesirable. This implies a relationship with the utility functions that drive decision theory.

Confirmation	Anticipated?	Effected Agent	Intensity Variables
(Dis)Confirmed	No	Self	Importance
(Dis)Confirmed	Yes	Self	Importance, previous probability, effort expended, degree realized
Unconfirmed	Yes	Self	Importance, probability
(Dis)Confirmed	No	Other	Importance to other, importance to self, extent deserved, extent (dis)liked

Table 3.1: Intensity factors for goals

This output can be made useful to such decision theory algorithms by creating a utility function that combines these emotions and their intensities into a single number representing the desirability of a course of action. The details of this are personality-dependent, as, for example, some individuals are extremely shame averse, and will avoid courses of action that lead to significant goal successes if they believe them to be morally reprehensible. Given such a utility function, various decision-making strategies then become applicable: score maximization, game theory, least regret, etc.

This paper describes a partially implemented system representing these ideas, using the OCC model to generate emotions. The scenario involves the planning of a terrorist bombing mission. The emotional outcomes of terrorist missions are particularly important to consider, as rarely are such attacks designed with force on force attrition in mind – it is precisely the emotional impact on the enemy and the general populace that makes the mission worth doing.

2. The OCC Model

As mentioned previously, the OCC model divides the concerns of an agent into goals, standards, and preferences.

2.1 Goals

Goals can take one of three forms: 1) active goals, which the agent can directly plan to make happen (I want to reload my rifle); 2) interest goals, which are states of the world that the agent would like to become reality but generally has no say in (I want important missions); and 3) replenishment goals, which periodically spawn active goals based on time since last fulfillment (I do not want to starve).

2.1.1 Active goals

Active goals are those states of the world that an agent is currently engaged in and attempts to bring about through direct manipulation of the environment. These typically manifest as the individual steps of a plan, and can be tightly integrated with a planner as demonstrated by Gratch [4]. An active goal succeeds when its post-conditions evaluate to true, and fails when the negations of its preconditions are confirmed to be true. The importance of an active goal can be modeled as inversely proportional to the number of acceptable alternative means of accomplishing the same step in the plan.

2.1.2 Interest goals

Interest goals differ slightly in that in general agents cannot take direct action to accomplish them. These become particularly important in game theoretic decision making, as an important interest goal of one agent may be entirely thwartable by the actions of an opponent.

It is not clear exactly how interest goals come to be held by an agent, and for this reason they are implemented as static parts of an agent's goal hierarchy with importance values set by the agent author. While this does oversimplify their role, there is likely a complex social and psychological process involved in the creation and maintenance of interest goals, and doing justice to this is beyond the scope of this current article.

In order to determine the intensities of emotions pertaining to the success or failure of goals, the OCC model uses several variables depending upon the context of the situation. Specifically, the variables used depend on whether the event is confirmed, disconfirmed, or unconfirmed, whether the event was anticipated, and whether it happened to the agent itself or someone else. Table 3.1 shows the intensity variables associated with each permutation of these variables.

Unfortunately, it is far from clear how some of these can be computed. For this reason, the system as implemented to date tracks only the importance, probability, and temporal proximity of goals. The variables pertaining to how one agent feels about another are considered relationship parameters, and will be discussed later. Degree deserved, effort expended, and degree of realization are left for future research.

2.1.3 Replenishment goals

Replenishment goals are essentially recurring active goals whose success or failure is a function of how long it has been since they were last fulfilled. As implemented, for some time after fulfillment they are considered to have succeeded. After this time they are considered unaffected

until, when a goal-specific amount of time has elapsed, they are considered to be failing.

2.1.4 Goal-based emotions

Under the OCC model, unanticipated confirmed goal successes and failures for one's self generate joy and distress, where anticipated goal effects in an unconfirmed state generate hope and fear. In a confirmed state, hope and fear will turn to satisfaction or disappointment, respectively, and in the disconfirmed state fear and hope become relief and, for lack of a better term, fears-confirmed. When evaluating how the goals of others have been effected, goal successes will generate happy-for or resentment, and goal failures will generate gloating or pity, depending on whether the agent in question is liked or disliked by the agent experiencing the emotion.

2.2 Standards

Standards are not unlike interest goals in that they are passive in nature. However, since they represent how people should behave, they are triggered not only when something relevant happens to the agent, but when something relevant happens to anyone. We are affected by reading accounts of ancient warfare practices not because these still can threaten us (or anyone we care about) in any tangible way, but because they often differ so greatly from what we consider acceptable.

2.2.1 Standards-based emotions

Standards are responsible for what the OCC model terms "attribution emotions". When responsibility for an action is attributed to one's self, pride or shame will result. When attributed to an external agent, these turn to admiration or reproach.

The intensities of standards-based emotions are effected by three primary factors. The degree of judged praiseworthiness or blameworthiness is the first, and is implemented as the result of the significance function for an effected standard.

The strength of the cognitive unit between the emoting agent and the agent performing the action determines the degree to which one will feel, for instance, shame as opposed to reproach for the blameworthy actions of another person. It is often the case that one will indeed feel a self-focused emotion about, for instance, the actions of one's country, even though that individual strictly had nothing to do with that particular action.

The third intensity factor involves deviations from role-based expectations. This captures the idea that we generally do not develop intense feelings about things we expect of people. As there is not yet a system in place for managing roles and the development of beliefs about another agent's concern structures, this intensity factor is not is not yet implemented.

2.3 Preferences

Lastly, preferences track the likes and dislikes of an agent. While typically pertaining to objects (I dislike broccoli), it is important particularly in military scenarios to note that it is entirely possible to view another agent as an object. This has the side effect of making them not subject to standards, and consequently an agent will not feel standards-based emotions about anything done to the objectified agent. This includes, for example, the shame normally felt for inflicting needless harm on another person.

2.3.1 Preference-based emotions

Emotions resulting from effects upon preferences come only in two varieties under the OCC model: liking and disliking.

Preference-based emotions have only two intensity factors. The degree to which an object is considered appealing or unappealing is modeled again using the significance function. As advertisers are well aware, familiarity with an object breeds a tendency to express a preference for it, and as the second intensity factor, this amplifies the intensity of liking and dampens the intensity of disliking.

2.4 Representation of Goals, Standards, and Preferences

In this implementation, goals, standards, and preferences, collectively referred to here as concerns, are arranged hierarchically, with parent nodes being that which motivates their children. A goal to write a paper may have as children finding a topic, doing research, opening a word processor, and typing. Each concern contains a fulfillment condition and a thwarting condition, indicating when it has been satisfied and when it has become impossible, and two importance values indicating the degree to which its success or failure directly causes the success or failure of its parent. Importance values range from 0 to 1, with 1 indicating that if the child succeeds/fails the parent succeeds/fails totally, and 0 indicating that the success/failure of the child is irrelevant to the parent. In this example, finding a topic is critical to writing the paper, giving it a failure importance of 1.0, but finding a topic is only a small part of finishing the entire paper, giving it a substantially smaller success importance. Figure 2.1 shows this graphically. The number before the slash is the success importance, and the number after the slash is the failure importance.

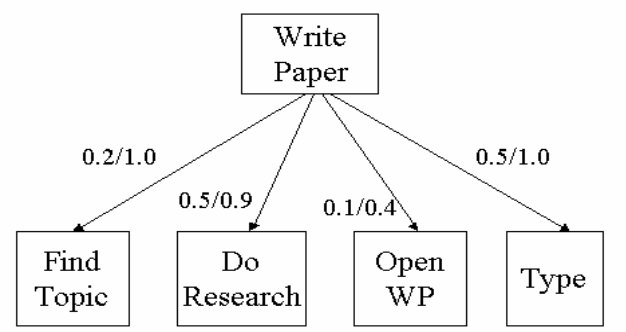

Figure 2.1: Hierarchical goal structure

Fulfillment and thwarting conditions are expressed using predicate logic with functions and relations defined in arbitrary Python code that draws upon the agent's knowledge of the world. Each logical statement is quantified over two variables indicating who is the agent performing the action and who is the direct object of this action. If a particular goal has been effected, a list of time intervals when the statement is believed to be true is returned along with confidence values (interpreted as probabilities) and variable bindings. Consequently, four pieces of information are conveyed to the agent: 1) whether this concern has been effected in the past or may be effected in the future, 2) how confident the agent is that this is the case, 3) who is responsible for this occurring, and 4) who was affected by this event. The structure of this information closely resembles that which is conveyed in the construal frames of Elliot [2].

As currently implemented, the functions used by these predicates are simple table lookups with values set by the scenario designer. When attached to a more complete agent model, however, they will draw upon the current beliefs of the agent. A number of important issues are hidden here, particularly credit/blame assignment and determining the probability and temporal proximity of future events.

2.5 Relationships

One of the most important functions of emotions is to regulate our behavior in social situations. As such, agents must represent not only their own concerns, but also those of the other agents they know. The OCC model uses four parameters involving how agents feel about one another, again dependent upon the type of emotion being generated. For each pair of agents (X, Y), the following are defined by the model: 1) the degree to which X likes Y; 2) the degree to which X dislikes Y; 3) the degree to which X has formed a cognitive unit with Y; and 4) the degree to which X is familiar with Y. Two additional parameters were added for implementation purposes, to be used in determining the intensities of standards-based and preference-based emotions, respectively: 1) the degree to which X views Y as an agent; 2) the degree to which X views Y as an object.

2.6 Calculating Intensities

At each event, all agents evaluate their goals, standards, and preferences, as well as those of the other agents they know. A success or failure is given a significance value by multiplying importance values up the hierarchy. The significance of a concern c is determined by the equation:

Equation 2.1

$$s_o(c) = \begin{cases} 1 \text{ if c is the root} \\ I_o(c, p(c)) * s_o(p(c)) \text{ otherwise} \end{cases}$$

where o is the outcome for a concern c (either success or failure), $s_o(c)$ is the significance of c for outcome o, I_o (c1, c2) is the importance under o of c1 to c2, and p(c) is the parent of c.

This significance value is then used in an emotion-specific equation along with other intensity factors discussed previously to determine the intensity of a specific emotion. These equations are given in Appendix A.

3. Linking Emotion to Utility

Appraisal models are consistent in their reliance upon a set of agent concerns, but for the most part give no advice about how to determine what these concerns might be for a fully developed agent. At the highest level, the works of Maslow [5] are relevant, pointing out five basic motives from which all others are derived: 1) physiological needs; 2) safety needs; 3) social belonging; 4) personal esteem; and 5) self-actualization needs. Though we reject his seriality premise, Maslow's work has been empirically shown to be descriptive of individuals across cultures, age groups, and generations, and is consequently a rich source of high-level goals, standards, and preferences.

In terms of creating reusable models of emotive agents, what is needed is a rich hierarchy of goals, standards and preferences for each type of agent. A good example of such a rich hierarchy for terrorists may be found in Weaver & Silverman [10]. That work shows a hierarchy of cases that differs across terrorists who come from different organizations. Further it shows how to devise hierarchies for new groups as a function of their political setting, ideology, campaign & mission aims, operational objectives, and so on.

At still lower levels, concerns vary widely among individuals. To develop a complete model of what an

agent cares about, we must probe deeper into who we are modeling. Upbringing, personal history, and individual quirks can all significantly effect what goals, standards, and preferences an agent is likely to hold. Two terrorists even from the same group will tend to have differences in their care, as will any two soldiers. However, we may not care to model such fine-gained differences. Also, several effects in clandestine terrorist cells tend to drive them to be consensual (e.g., being isolated from others and needing to belong to the cell, as well as the well known "risky shift" effect).

Finally, it should be noted that when describing a group of people after extensive study, the language used by authors often directly translates into a description of their common goals, standards, and preferences, and how they differ from other groups. From this we hope to derive a reusable database of archetypes, from which we can provide agent authors a template to use in instantiating members of an organization.

Even given such a system, we must account for the effects of one more dimension of individual differences. Despite similar emotional outcomes, different people will often still choose different alternatives: some are pleasure seekers, some are tremendously averse to distress, and still others will endure great pain as their goals fail in order to uphold their standards. We have chosen to represent this observation by taking into account the "Big 5" personality traits [7].

According to this taxonomy, the personality of an individual can be parameterized into five dimensions: surgency, agreeableness, conscientiousness, emotional stability, and openness to new experiences. Each of these dimensions is implemented as scored from 0 to 1, and acts as a weight upon certain emotions when combining into utility.

The term *surgency* refers to the degree to which agents are proactive in achieving their goals. Individuals strongly exhibiting this trait consider advancement of their own goals to be of paramount importance, potentially at the expense of failing standards and preferences, or negative emotional outcomes for others. A surgent individual will not think twice given an opportunity to wade through raw sewage for a chance to surprise an unprepared enemy. As implemented, this trait weights the importance of and joy, satisfaction, relief, and liking.

The second factor in the taxonomy is *agreeableness*. Agreeable individuals are strongly concerned about the goals of others, and will often suppress their own to see them satisfied. An agent dominated by this trait will often betray his instincts to follow orders. This trait weights the contribution to utility of gloating, pity, happy-for, and resentment.

Conscientiousness is the third trait of the Big 5, and measures the degree to which agents consider the full ramifications of their actions before taking them. Those strongly exhibiting this trait avoid courses of actions leading to negative consequences, even if accompanied by substantial positive ones. Such agents are unlikely to choose courses of action considered dishonorable, or risky, often at the expense of opportunities to achieve goal successes. This trait is used as a weight for distress, fears-confirmed, disappointment, disliking, pride, shame, admiration, and reproach.

The fourth factor is *Emotional Stability*, which, for decision-making purposes, governs the degree to which an agent is willing to endure pain along the path to goal achievement. An emotionally stable person, despite moral and other objections to a course of action, may still choose it if under the impression that it will have a significantly positive effect on things later on. This term weights the importance of immediate gratification by recursively adding the utilities of an imagined successor states.

Openness to Experience captures the observation that occasionally agents will choose an emotionally-neutral, not previously explored course of action over one proven to provide some degree of gratification. This is accomplished by having this parameter act as a negative weight upon the most intense emotion generated by a course of action.

Given a full model of the concerns of an agent, an emotion model that combines these with events to create feelings, and these five personality factors, we must still combine these into a single number in order to utilize the wealth of pre-existing decision theory algorithms in existence. We use the following equation to convert emotion intensities to utility using personality:

Equation 3.1

$$\begin{aligned} U(c) = {} & P_s * (E_{\text{joy}} + E_{\text{satisfaction}} + E_{\text{relief}} + E_{\text{liking}}) + \\ & P_a * (E_{\text{happyfor}} + E_{\text{resentment}} - E_{\text{gloating}} - E_{\text{pity}}) - \\ & P_c * (E_{\text{distress}} + E_{\text{fearsconfirmed}} + E_{\text{disappointment}} + E_{\text{disliking}}) + \\ & P_c * [(E_{\text{pride}} - E_{\text{shame}}) + (E_{\text{admiration}} - E_{\text{reproach}})] - \\ & P_e * \max(E) + \\ & P_m * U(\text{successor}(c)) \end{aligned}$$

where $U(c)$ is the utility of course of action c; P_s, P_a, P_c, P_e, and P_m are the personality dimensions surgency, agreeableness, conscientiousness, openness to experience, and emotional stability, respectively; and E_x is the

maximum intensity of emotion x (I_x) over all possible concern effects times the perceived probability of this outcome actually occurring.

As currently implemented and in the following example, the system only determines courses of action one step ahead, and consequently the emotional stability term has no effect.

3.3 Example Scenario

To illustrate the processes described above, consider the problem of mission planning and executing from the point of three different terrorists, all of whom share a common enemy. Terrorists A and B also share a common set of goals, motivated by a religious conflict of interests with their enemy. However, Terrorist B believes passionately that sacred landmarks, even those of conflicting religions, should not be desecrated, where Terrorist A holds no such standard. Terrorist C, while sharing many basic goals with A and B, has been driven to commit an act of terrorism based on a difference in political ideologies rather than religion. Specifically, Terrorist C is a communist striking against a capitalist regime.

Weaver, Silverman, et al.[10], present a framework for semi-automatically generating the utility structure of the terrorist groups, such as A, B, and C. This utility structure emphasizes the importance of missions to the campaign, targets to missions, and operational details, and how all this effects the population. It is particularly important for terrorists to carefully consider how their actions will effect their relationship with the surrounding population. Additionally, an action means nothing if the enemy is unaffected. Therefore, each terrorist is concerned about the potential outcomes for himself, his enemy, and an aggregate agent representing the general populace. Specifically, each terrorist holds a goal that succeeds when (and to the extent that) the populace is positively affected, and fails when the populace is negatively affected. Since this conflict is essentially a zero-sum game between the terrorist and his enemy, game theory is used to model the decision-making process. That is, the Weaver, Silverman, et al. framework is meant to be used offline to generalize structural differences between groups (and individual agents). Here we explain a dynamic framework for using their need structures to process emotions and personality differences.

To simplify the example, we will assume for now that all three terrorists share the same scores for each personality trait. Let $P_s=P_a=P_c=P_e=P_m=0.5$.

Consider the target selection process. The three terrorists are aware of five potential targets: a bank, a sports arena, a religious landmark, a government building, and a military outpost. The relevant goals and standards of Terrorist A are shown in Figures 3.1 and 3.2, respectively.

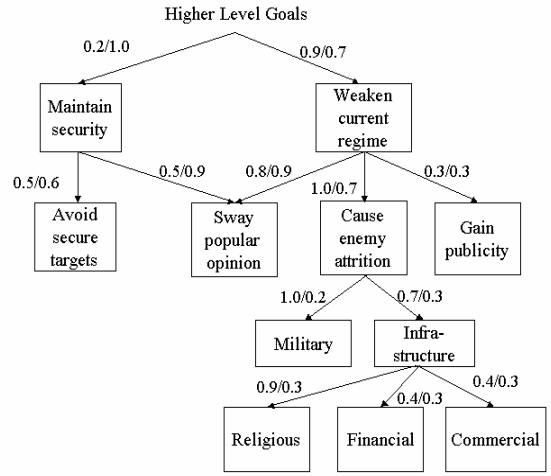

Figure 3.1: Goals for Terrorist A

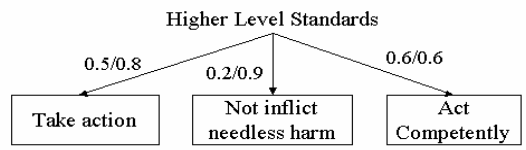

Figure 3.2: Standards for Terrorist A

To select a target, Terrorist A will first examine the utility to himself of attacking each, beginning with the bank. On the positive side, as a high-profile, highly secured target, a successful mission against it will gain significant publicity for the organization in addition to having an impact on the financial infrastructure of the current regime. Furthermore, Terrorist A has no moral objection to this course of action, and will indeed feel some pride upon successfully striking this target since he has taken action to end an undesirable situation. However, due to its high security, his goal to avoid getting caught, and in turn the security of the entire organization is threatened.

To determine whether popular opinion will be positively or negatively effected, it is necessary to evaluate the emotional outcome of this situation for what the terrorist believes to be the concerns of the general populace. The goals and standards of these are shown in Figures 3.3 and 3.4, respectively. In this case, a mixture of joy, reproach, admiration, and distress results, causing the goal to sway popular opinion to be threatened. Consequently, as a target the bank has the potential to create high levels of joy (attrition and publicity), distress (getting caught and losing crowd support), and lower levels of pride. Given equal weightings from personality factors, these are combined via equation 8.1 to produce a utility of x. A partial calculation, showing the contribution made to utility by

the potential event-based emotions generated by attacking a secure target, is shown in Table 3.1.

Emotion	Intensity	Prob.	Partial U
Distress	0.6	0.9	-0.54
Joy	0.1	0.1	0.01

Table 3.1

Thus the high probability of the failure of this goal will cause a substantial (-0.53) lowering of the utility of this alternative in all but the least conscientious of agents.

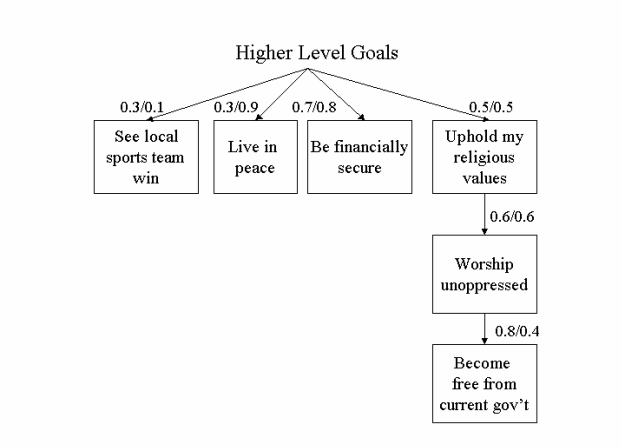

Figure 3.3: Terrorist's View of Populace's Goals

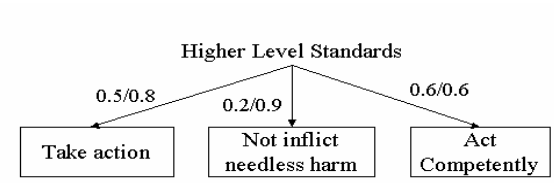

Figure 3.4: Terrorist's View of Populace's Standards

A similar process is undergone to determine the utility of the other targets, as shown in Table 3.2. Note that given the importance of religion to this terrorist, attacking a religious landmark is considered most attractive in terms of goal achievement. Furthermore, unlike Terrorist B, Terrorist A has no standard indicating that this is unacceptable behavior.

Since we have chosen to represent the decision-making process using game theory, we must also now determine the utility of each of the possible terrorist courses of action for the enemy. A gratifying, successful attack on a target is not nearly as attractive if it creates an opportunity for the enemy to eliminate his organization entirely or turn the populace further against him. Additionally, in the absence of an opportunity to directly achieve his goals, he may be able to put the enemy in a situation in which they do it for him. Consider again the utility of a bank bombing, this time from the point of view of the regime in power, whose leader's goals and standards are shown in partially in Figures 3.5 and 3.6.

Target	Utility to Terrorist	Primary Contributions
Bank	-0.27	Security threat (-), populace reaction (-)
Arena	-0.56	Populace reaction (-), publicity (+)
Government building	-0.12	Security threat (-)
Religious landmark	0.32	Attrition (+), populace reaction (-)
Military Outpost	-0.12	Security threat (-)

Table 3.2

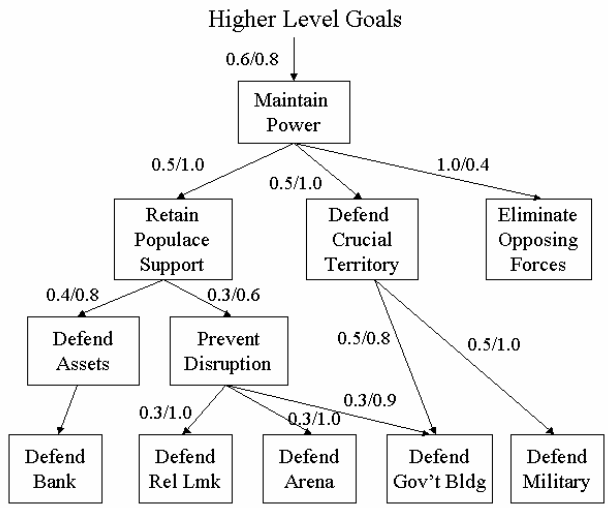

Figure 3.5: Enemy leader's goals

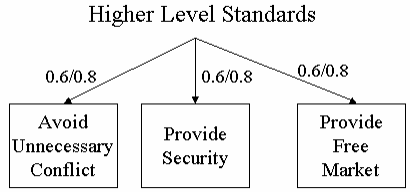

Figure 3.6: Enemy leader's standards

In short, the enemy would love to see a terrorist target its bank. Given the security measures in place, a number of goal successes are highly probable: threat elimination, demonstrating the security of the populace's interests (and

in turn securing valuable support from them), etc. The successful defense of this territory will also generate pride, and consequently, the utility of this course of action to the enemy will be extremely high, making it less attractive to the terrorist than originally estimated. A similar process is undergone for each other target, with results shown in Table 3.3. Note again that due to the importance of the religious institution to the populace and its lack of security, the utility of having this building targeted is extremely low to the enemy. Given that it is already highly attractive to this terrorist, it should come as no surprise which target will be selected by Terrorist A.

Target	**Utility to Enemy**	**Primary Contributions**
Bank	0.65	Terrorist attrition (+), populace reaction (+), positive press (+)
Arena	-0.29	Shame (-), populace reaction (-)
Government building	0.51	Terrorist attrition (+), pride (+), positive press (+)
Religious landmark	-0.48	Populace reaction (-), shame (-)
Military Outpost	1.1	Terrorist attrition (+), pride (+)

Table 3.3

We now turn to the decision-making process of Terrorist B. Since they have the same goals and nearly the same standards, the utility of all targets excepting one is unchanged. However, due to his strict objection to destroying religious ground, even that of a religion to which he is violently opposed, a significant amount of shame would be induced by selecting this alternative, lowering its utility enough that Terrorist B will choose Terrorist A's second choice, the sports arena.

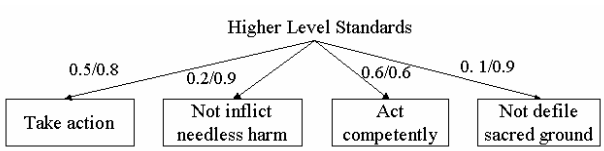

Figure 3.7: Terrorist B's standards

Terrorist C is motivated by substantially different concerns from A or B. Since his primary motivation is political/economic in nature, targets involving commerce begin with higher importance values attached to their success in the goal structure. Consequently, they generate a higher potential for joy. However, as this terrorist has the same resources as the other two, the successful bombing of high security areas is still quite unlikely, and it is still highly undesirable to get caught. Terrorist C then faces will choose the sports arena as well, citing its symbolic value as a center of commerce in addition to its vulnerability.

4. Conclusions and Next Steps

This paper has explained a framework for integrating cognitive appraisal (emotion models) and personality theory into agent decision-making. This framework is based on the OCC model, but that model alone only generates emotional state. It is unable to provide agent decision guidance. This paper's contribution is to provide one way to extend cognitive appraisal in general, and the OCC model, in particular, into a mechanism for choosing between alternative decisions and courses of action. We do that mathematically by trying each of the 22 OCC emotions in a principled way to one of the "Big 5" personality factors and by using that in a utility calculation equation. Thus, any event, agent action, or object precipitates a utility based on emotional intensity and on personality weights. These utilities in turn are what determine decision options in the classical game-theoretic approach.

Reducing emotion and personality to utility calculation may be elegant, and it may even be a computational advance for the agent field, however, that does not alleviate us from a number of validity concerns. We have raised some of these concerns earlier such as the validity of the OCC model (and its choice of 22 emotions), ignoring nuances and subtleties of emotions and personal reaction at the fine-grained level, and the lack of empirical support for only 5 factors in the Big 5 model. We have also mentioned our assertion that validity concerns are partially mitigated by the higher granularity that most models address. However, one could more explicitly address such concerns via a Monte Carlo simulation covering variations one might expect to arise at the fine-grained levels. We have not attempted such an approach.

The system described here was designed to be integrated into a larger agent model. As such, a number of important features are handled crudely, as they are simply placeholders for the deeper functionality offered by a more complete model. Probability assessment and credit/blame assignment are primary among these. Also, we are currently involved in an effort to integrate this system as part of a larger agent model into JSAF.

To represent the dynamic nature of active goals and their relationship with interest and replenishment goals, it would be beneficial to tightly integrate the decision-making process with a planner, and feed the results back into the goal structure. This will allow not only make the model more complete, but may also begin to implement the problem-focused coping mechanisms described by Lazarus [5]. By not explicitly modeling emotion decay as a function of time, our agents will be forced to make decisions taking into account not only the emotional outcome of their actions, but also whether or not they help to solve some pre-existing problem. If a terrorist runs out of gasoline on the drive to his target, he has no choice but to first eliminate the associated distress – the potential joy of mission accomplishment cannot be attained with no way of getting to the target, so every course of action aside from finding a gas station will result in nothing but persistent distress and likely shame.

In the event that an emotion cannot be eliminated by planning a way to make it stop failing, it will be necessary to model what Lazarus terms emotion-focused coping, which can be modeled by shifting importance values in the concern structures so that the effected concern becomes insignificant enough to stop causing an emotion. While far from straightforward to implement, this approach should yield a powerful method of creating agents whose concerns change with experience. We will thus have agents who become demoralized, complacent, obsessed, or bored, among other things.

Agents currently have complete knowledge of one another's concern structures. This is far from realistic, and could have dramatic effects if a model of how people acquire such information from one another is implemented. This is primarily a recognition problem, involving taking raw data about some history of actions taken by an agent, their outcome, and what emotional reactions were observed, and attempting to assemble this into a model of what motivated these actions.

As currently represented, relationships are also static and determined in a rather ad hoc fashion. It may be possible to derive these parameters directly from evaluating the concerns of one agent against what she believes to be the concerns of another. Goal compatibility likely correlates with liking an individual. Having many goals effected in the same way by the exact same events likely contributes to a substantial cognitive unit between two parties. Familiarity may simply be the extent to which one believes his model of another to be complete. By deriving these parameters in this way from the beliefs of agents about others, we obtain a dynamically evolving relationship.

This framework lends itself naturally towards an exploration of implementing the thoughts of Damasio, whereby decision-making will be done in not only a more naturalistic way, but also using potentially far less computation time.

Appendix A: Intensity Equations

$$I_{\text{happyfor}} = S(sg) * R_l$$

$$I_{\text{gloating}} = S(sg) * R_d$$

$$I_{\text{resentment}} = S(fg) * R_d$$

$$I_{\text{pity}} = S(fg) * R_l$$

$$I_{\text{hope}} = S(sg) * P(sg) * N$$

$$I_{\text{fear}} = S(fg) * P(sg) * N$$

$$I_{\text{satisfaction}} = \text{Previous } I_{\text{hope}} * (1 - N)$$

$$I_{\text{fearsconfirmed}} = \text{Previous } I_{fear} * (1 - N)$$

$$I_{\text{relief}} = \text{Previous } I_{\text{fear}} * (1 - N)$$

$$I_{\text{disappointment}} = \text{Previous } I_{\text{hope}} * (1 - N)$$

$$I_{\text{joy}} = S(fg) * O * (1 - N)$$

$$I_{\text{distress}} = S(fg) * O * (1 - N)$$

$$I_{\text{pride}} = S(ss) * R_c$$

$$I_{\text{shame}} = S(fs) * R_c$$

$$I_{\text{admiration}} = S(ss) * [(1 - R_c) + R_a] / 2$$

$$I_{\text{reproach}} = S(fs) * [(1 - R_c) + R_a] / 2$$

$$I_{\text{gratification}} = \begin{cases} 0 \text{ if } \text{abs}(I_{\text{pride}} - I_{\text{joy}}) > .2 \\ \max(I_{pride}, I_{\text{joy}}) \text{ otherwise} \end{cases}$$

$$I_{\text{remorse}} = \begin{cases} 0 \text{ if } \text{abs}(I_{\text{shame}} - I_{\text{distress}}) > .2 \\ \max(I_{\text{shame}}, I_{\text{distress}}) \text{ otherwise} \end{cases}$$

$$I_{\text{anger}} = \begin{cases} 0 \text{ if } \text{abs}(I_{\text{reproach}} - I_{\text{distress}}) > .2 \\ \max(I_{\text{reproach}}, I_{\text{distress}}) \text{ otherwise} \end{cases}$$

$$I_{\text{gratitude}} = \begin{cases} 0 \text{ if } \text{abs}(I_{\text{admiration}} - I_{\text{joy}}) > .2 \\ \max(I_{\text{admiration}}, I_{\text{joy}}) \text{ otherwise} \end{cases}$$

$$I_{\text{liking}} = S(sp) * R_o * R_f$$

$$I_{\text{disliking}} = S(fp) * R_o * (1 - R_f)$$

$$N = \begin{cases} 1 \text{ if the time interval in question begins after now} \\ 0 \text{ otherwise} \end{cases}$$

$$O = \begin{cases} 1 \text{ if we have previously had a} \\ \quad \text{prospect-based emotion about this goal} \\ 0 \text{ otherwise} \end{cases}$$

where sg is a succeeded goal; fg is a failed goal; ss is a succeeded standard; fs is a failed standard; sp is a succeeded preference; fp is a failed preference; R_l, R_d, R_c, R_f, R_a, and R_o are the liking, disliking, cognitive unit, familiarity, agent, and object relationship parameters, respectively; P is a function determining the probability of the success or failure in question occurring; and S is the significance function given previously.

References

[1] Silverman, BG, Might, R., et al., "Toward A Human Behavior Models Anthology for Synthetic Agent Development", *10th CGF Proceedings*, 2001.

[2] Damasio, Antonio: *Descartes Error: Emotion Reason, and the Human Brain*, Avon Books, New York, 1994.

[3] Elliot, Clarke: "The Affective Reasoner: A process model of emotions in a multi-agent system", Doctoral Dissertation, Northwestern University, Evanston Illinois, 1992.

[4] Gratch, Jonathan: "Modeling the Interplay Between Emotion and Decision-Making", *Proceedings of the 9th CGF*, 2000.

[5] Lazarus, Richard: *Emotion and Adaptation*, Oxford University Press, Oxford, 1991.

[6] Maslow, Abraham, *Motivation and Personality*, 2nd ed., Harper & Row, 1970

[7] Ortony, Andrew, Gerald L. Clore and Allan Collins: *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge, 1988.

[8] Revelle, William. *The Personality Project*. http://pmc.psych.nwu.edu/personality.html

[9] Roseman, Ira, Martin S. Spindel, and Paul E. Jose: "Appraisals of Emotion-Eliciting Events: Testing a Theory of Discrete Emotions."

[10] Weaver, R., Silverman, BG, et al., "Modeling and Simulating Terrorist Decision-making", *10th CGF Proceedings*, 2001.