

# Adaptive Decision Models for Simulating Co-evolving Adversarial Behavior

Dr. Gary L. Klein

The MITRE Corporation  
7515 Colshire Drive, M/S H205, McLean, VA 20171  
gklein@mitre.org

## Abstract

An adaptive agent-based simulation modeling technology has been developed that allows us to build, for example, simulated decision makers representing defenders and attackers of a computer system engaged in cyberwarfare in their simulated microworld. The adaptive adversaries co-evolve: attackers evolve new attack patterns and overcome cyber defenses, and defenders subsequently evolve new defensive patterns to the attacks. When we run these adaptive decision-maker models, we see what looks like human adversarial behavior. These simulated attackers learn to time their attacks just as real-world hackers do with virus attacks. Simulated defenders soon catch on and resynchronize their defenses to match the timing of these attacks. This adaptive simulation modeling can automatically discover new behaviors beyond those that were initially built into the models, providing a more realistic simulation of intelligent behavior. Such models provide both an opportunity to discover novel adversarial behavior, and a testbed for other adversary course of action prediction models.

## Adaptive Decision Agents

An adaptive agent-based simulation modeling technology has been developed that allows us to build simulated decision makers. The adaptive ability of these agents is achieved by providing them with methods that allow them to perceive the results of their actions and then to modify their behaviors to improve their performance in achieving their goal-state. As an agent interacts with its simulated environment it must solve three decision making problems: determining which inputs to use; how to combine them; and how to weight each input in order to determine the appropriate course of action.

The agents' threefold decision problem was simulated in terms of a statistical multiple regression model. Multiple regression models have been shown in the psychological literature to well capture human decision making behavior even if they do not represent the underlying cognitive processes leading to that behavior. This distinction between behavior and underlying process will provide the essence of the testbed concepts presented later.

Mathematically, the regression model is a polynomial and can include non-linear terms represented by the inclu-

sion of the  $A^2$  term in an example regression model form below:

$$Y = b_0 + b_1A^2 + b_2B + b_3CB$$

In our new adaptive implementation of this technique, each decision agent in the simulation was actually composed of a "molecule" of agents as illustrated graphically in Figure 1. In the illustrated example the decision-agent is attempting to estimate an Airport Arrival Rate (AAR)

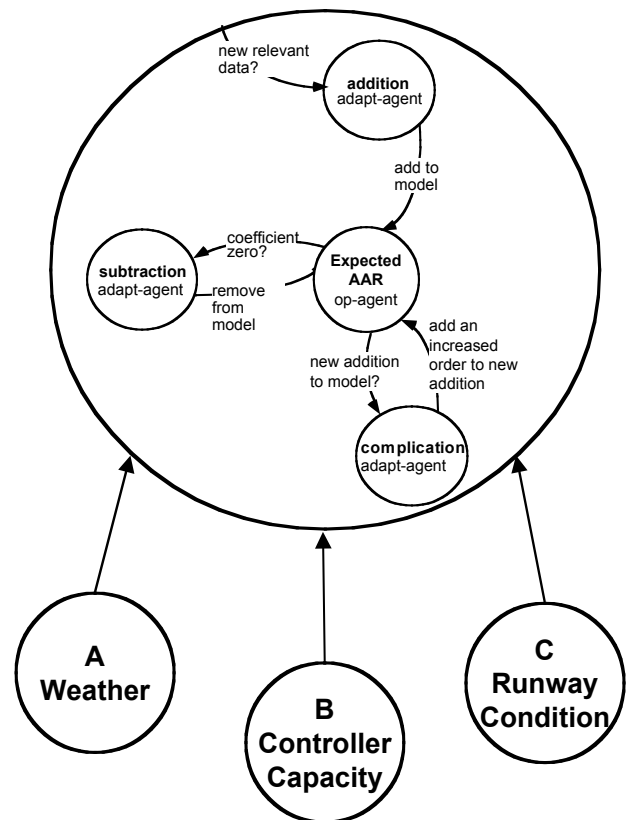


Figure 1. Graphical representation of decision agent architecture

based on Weather, Controller Capacity, and Runway Conditions. The core “op-agent” optimizes the regression weights ( $b_0 \dots b_n$ ) as feedback becomes available on actual AARs for comparison with its predictions.

To go beyond the mere calculation of optimal weights for an *a priori* set regression model, “adaptation-agents” that surround the op-agent in the illustration were developed. These agents use heuristic and statistical methods to determine whether to change the right-hand terms (A...C) in the model. When the environment changes (due to the evolving actions of other agents) or when the agent is exposed to new situations, even the best regression fit between an agent’s model and the available data could fall below an acceptability threshold. When that happens, the adaptation-agents modify the terms of regression model (adding or deleting input sources or trying a different combination of inputs). In this way, the agent is always exploring the space of possible models and adapting to deal with current contingencies.

As described above, these decision-agents engage in a form of “supervisory” learning. Where the feedback the agent would receive from its environment is of the same dimension as the values it is trying to predict (e.g., predicted AAR vs. actual AAR).

Applying a somewhat more involved method, “reinforcement” learning also has been implemented, where the feedback and agent outputs are on different dimensions. For example, a decision-agent may be trying to determine the number of airplanes to depart in order to maximize its net utility. Thus the decision-agent’s output would be in terms of departures while the feedback would be in terms of profits. By regressing past utilities (U) against past departure decisions and inputs into those decisions (the right-hand side of the model), optimal weights and terms can be determined as previously described. However, when a decision is to be made, a *predicted* U is not the output; instead a *target* value for U is provided as input along with the values of the other right-hand terms, *except for a value the departure term*. With all of the other values for the other terms of the model supplied, and the weights determined from historical data, the equation can be solved for the value of the departure term. This value is the output of such a decision agent. In this way, the agent learns the appropriate departure behavior with regard to local input conditions as a result of the reinforcement it receives in terms of net utility.

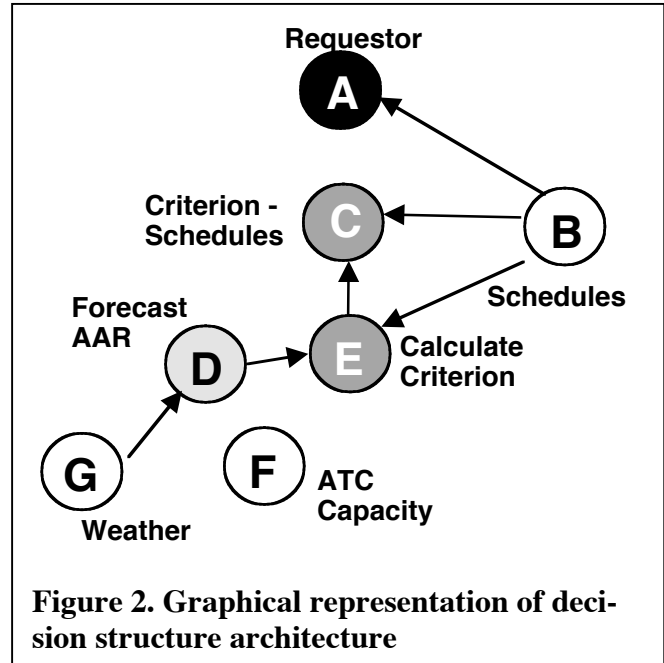
Through these kinds of supervisory and reinforcement learning the decision-agents are capable of adapting to go beyond their initial programming.

### Adaptive Decision Structures

The decision-agents and additional subsidiary agents can be organized into networks or *decision structures*. In these structures, the input values for terms in one agent’s model can come from the output of other agents. In this way, when one agent’s model is adapted, its links to other agents

will change and the emergent behavior of the decision structure will change as well.

An example of one such structure from a past research application is illustrated in Figure 2. *Requestor* is an example of a reinforcement learning agent. It is fed input values from *Schedules* which is a “sensor-agent” (as are *Weather* and *ATC Capacity*), which picks up raw data from its environment and pushes it to the agents to which it’s output is



linked. In Figure 2, *Schedules* also pushes data to *Criterion Schedules* and *Calculate Criterion* which are “integrator-agents”, which merely perform fixed calculations on their inputs to be passed to their output links. Note that in the illustration, the chain is currently broken because *Criterion Schedules’* output is not yet linked to any agent. In this example, the remaining agent, *Forecast AAR*, is a reinforcement learning agent.

A decision structure is defined by determining the deci-

	Decision	Candidate Sources	Feedback
A	Requestor	B C	Net Utility
B	Schedules	Data	
C	Criterion - Schedules	B E	
D	Forecast AAR	G F	Actual AAR
E	Calculate Criterion	B D	
F	ATC Capacity	Data	
G	Weather	Data	

Table 1. Example decision matrix fragment

sions to be made by the structure, the decisions' candidate sources of information, and the feedback by which decision-agents will adapt. An example decision matrix for the structure in Figure 2 is illustrated in Table 1. As represented in the table, *Criterion Schedules* is a candidate source for *Requestor*. *Requestor's* adaptation agents could therefore eventually add *Criterion Schedules* to *Requestor's* model and the output from *Criterion Schedules* would then be linked to *Requestor*.

### Sample Research

This powerful new simulation paradigm allows us to systematically examine how behaviors of *adaptive* individuals affect the *evolution* of the individuals and consequently of a society. If we simulate two societies with competing goals we can examine how they co-evolve. If we place adaptive agents on a simulated battlefield, then we have a simulated wargame where the parties are capable of adapting to their environment and to each other. This is how we developed the simulated cyberwargame described above. A decision structure of these agents may represent an organization, or a single mind of a decision maker.

Looking at it in another way, by enabling these agents to go beyond their original programming, we make them capable of extending our initial assumptions about how people behave in a given context. The agents can learn from and adapt to that context to best satisfy the goals we have given them.

Our first use of adaptive decision modeling (Klein & Antón, 1999) was to simulate how air traffic decision makers would perform under a number of traffic flow management operational policies ranging from no central traffic flow control to two forms of free-market control structures and strict central control.

It was reasonable to expect that rational decision makers would develop different strategies to deal with the different operational policies. And, there is solid empirical research on which to build *general* models of human decision making. But, we were dealing with hypothetical operational concepts, never tried in the field, and therefore there was no empirical data on which to develop *specific* models of decision making under each operational condition.

Adaptive decision modeling technology allowed us to do *exploratory* modeling. Our generic decision-making simulation models were placed in simulations of each condition, and allowed to explore automatically the strategies that best fit their conditions. Basically, the decision-agents competed in a repeated "coordination game" format, where each round of the game they vied for scarce landing spots at a capacity reduced airport. The experimental design is illustrated in Figure 3. One major advantage of using simulated decision makers was that we could create a control airline goal condition where we forced the air carriers to be altruistic and attempt to optimize system utility rather than their own utility. This showed that the informational differences across management conditions resulted in sig-

Airport Arrival Manager Response	Air Carrier Goal Conditions	
	Baseline: System Utility	Individual Utilities
No Arrival Management	NB System: 89.0	NI AC 1: 82.6 AC 2: 83.2 System: 77.9
Aggregate total requests	AB System: 90.0	AI AC 1: 96.2 AC 2: 83.5 System: 95.8
Suggest based on system utility forecasts	SB System: 88.5	SI AC 1: 87.3 AC 2: 82.5 System: 83.4
Constrain to allocations	CB System: 88.1	CI AC 1: 83.4 AC 2: 86.1 System: 78.2

**Figure 3. Experimental design and average utilities under each condition**

nificant performance differences only when self-interest was in play.

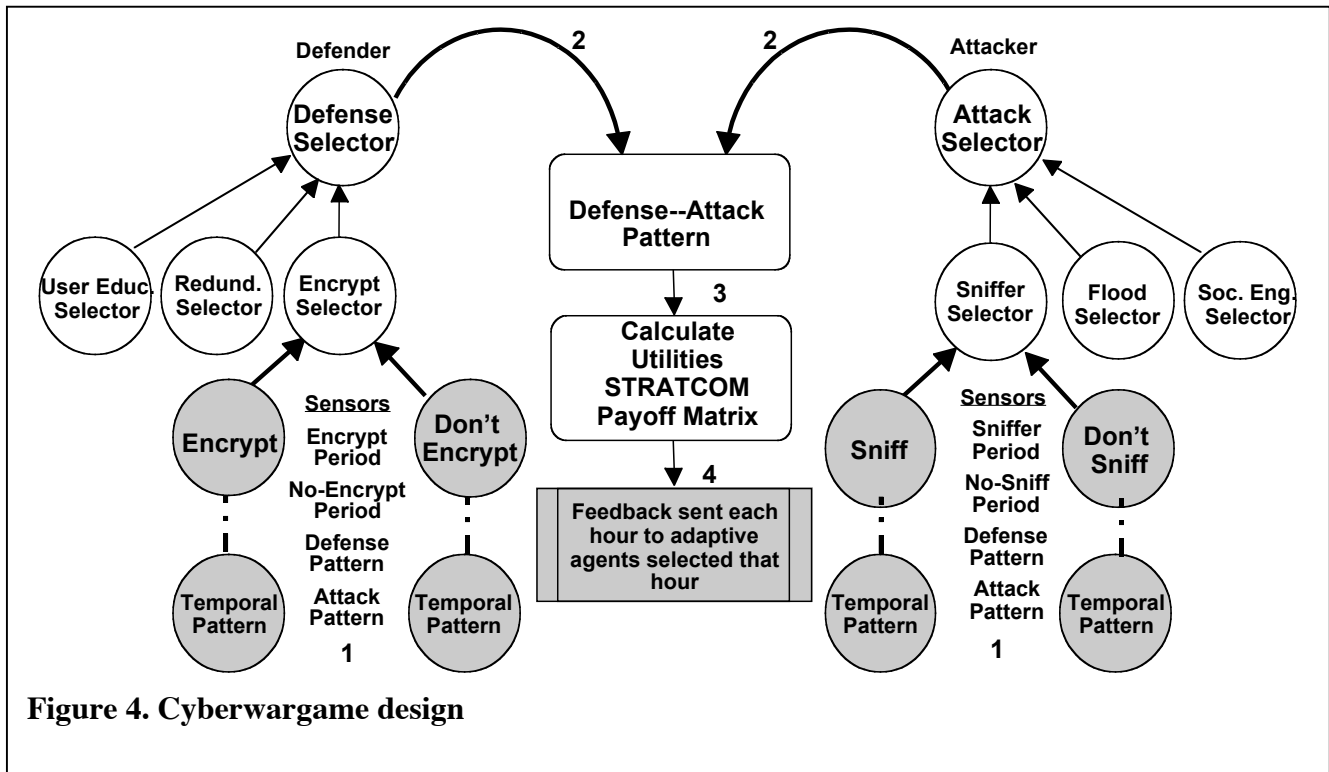
Under each of the simulated conditions the adaptive models did indeed evolve differently, and that resulted in different relationships evolving between the airlines and central flow management. With these simulated decision makers we could actually examine those differences in detail. For example compare the two different regression models that evolved for the two air carrier *departor* agents under one condition. For AC1, utility was only a function of the number of flights departed, and the number of its own scheduled flights. However, for AC2, it's utility function included:

- Number of flights departed
- Difference between a forecasted criterion and the allocation given it by arrival management
- 1/forecasted criterion
- AC2's schedule
- AC1's schedule/forecasted criterion

Across conditions, these different relationships resulted in different levels of system and individual performance, as can be seen in Figure 3.

Albeit in a simplified environment, these results and the nature of the evolved differences highlighted important factors that could affect the performance of traffic flow policies.

Because this was also a competitive situation, the simulated decision makers could optimize their performance only by either synchronizing their behavior (de facto cooperation) or by constantly changing their strategies as they seek to gain advantage over each other. In this controlled simulated environment, they were not allowed any direct



**Figure 4. Cyberwargame design**

communication (no collusion) with each other. Even so, in all but one experimental condition, our simulated decision makers learned to synchronize their behavior, and the pattern of synchronization changed over time. The results suggest that apparent cooperation can indeed evolve in a competitive situation without collusion.

The cyberwarfare model described at the beginning of this paper is our latest application of adaptive decision modeling. In a warfare situation, the co-evolving *adversarial relationship* between attacker and defender is the primary focus of the research. With an adaptive decision modeling experimental paradigm we can systematically manipulate the context of the warfare (initial defensive configuration, information available to each side, available resources, etc.) and evaluate the effect on the adversarial relationship. Figure 4 illustrates the basic design of one of the wargames.

This wargame is also a repeated coordination game. This one is a kind of sophisticated “scissors/paper/rock” game. Each adversary can choose none, one or some combination of three actions to take on each round of the game – resulting in 7 possible defensive decisions and a matching set of 7 offensive decisions. For each of an adversary’s three actions, there is a decision-agent predicting the utility of doing that action on this round, and a decision-agent predicting the utility of not doing it on this round. These action advocates are adaptive agents. Their predictions are output to an integrator-agent that selects whichever option has the highest predicted utility. This selection is then output to another integrator that combines the selection with those of the other selectors into a combination strategy. A

payoff matrix determines the utility each of the players receives based on how well the defender’s combination matches the attacker’s combination.

From this simple game and other variation upon it came some remarkable results. These simulated attackers learn to time their attacks just as real-world hackers do with virus attacks. Simulated defenders soon catch on and resynchronize their defenses to match the timing of these attacks. In fact, any persistent attacker or defender pattern was quickly learned and countered. Even more intriguing, over a number of variations on the game, the evolved counter-strategies on both sides were superior to a static game-theoretic mini-max solution.

In addition, one result of these cyberwar experiments duplicated an intriguing result in the air traffic experiments. In both cases some information needed to evolve a behavior is not needed to maintain the behavior. Perhaps even more intriguing is that without this eventually discarded information, the behavior may never evolve. For instance, if a game is played again, with fresh simulated adversaries, but an adversary is limited from the beginning only to the information used in a previous final evolved regression model, the fresh adversary never evolves the appropriate models or behavior. Adaptation must pass through certain intermediate stages for the behavior to reach the end state.

## Conclusions

This adaptive decision modeling simulation technology can be used to extend our knowledge and models of human

behavior and the behavior of complex adaptive systems, such as social systems, competitive markets, or adversarial behavior in combat. Adaptive agent technology provides automated discovery of behavioral models that would be impractical or impossible to search for with real-world experiments because of the size of the possible model space. Discovering these models in simulation will point us to significant variables and critical values of those variables that need to be examined under real-world conditions. Leveraging these discoveries will allow us to then use our real-world experiments more effectively to test, verify, and extend the knowledge made possible by these new simulation models.

In addition, there is another possible use for these adaptive decision models. As noted earlier, decades of research has shown that these models though not representative of human psychological processes can still model human decision behavior well.

This raises the intriguing question as to whether they might provide an interesting testbed for other approaches being used for modeling user/adversarial intent. It is the purpose of such intent models to predict the behavior of human decision makers. Then, can such models predict the adaptive behaviors of these simulated decision makers?

This could be an efficient way of testing and improving the usefulness of such intent models. Most such intent models for predicting behavior are based on qualitatively different models of behavior than the netted multiple regression models these adaptive agents use for generating human-like behavior. Where would any two such models coincide in prediction and result? When they do coincide or when they do differ, a decision-agent testbed would let one examine in detail the causes of the resulting behavior that led to its variance with prediction. Such experimentation could once again point us to significant variables and critical values of those variables that need to be examined under real-world conditions.

## References

Klein, G. L., & Antón, P. S. 1999. A Simulation Study of Adaptation in Traffic Management Decision Making under Free Scheduling Flight Operations. *Air Traffic Quarterly* 7(2):77-108.