Validating a Hybrid Cognitive-System Dynamics Model of Team Interaction

Asmeret B. Bier and Michael L. Bernard

Cyber Engineering Research Institute
Sandia National Laboratories
Albuquerque, NM 87185-1327, USA

ABSTRACT

Computational models of human behavior can lead to important insights regarding how people interact with each other and with their environments; but validation of these models is difficult and data is generally hard to collect. Better validation strategies could help to make these models more strongly justified and thus more useful. This paper describes an effort to study cooperation between teams in cyber security training exercises by building a model that captures the interactions between them. Real-world exercises provide a useful source of validation data and can serve to help calibrate the model. In this study we simulated two cyber scenarios where the primary difference was the intensity of cyber attacks experienced by two organizations. The model simulated the potential outcomes and decision-making processes involved in cooperative cyber security agreements designed to reduce redundant work. Insights learned from the model are intended to improve future versions of the cyber exercise. We also used validation data and insights from the exercise model to create and justify decision-making strategies in a model of a real-world counterpart to the situation exercises: information sharing programs for cyber defense.

Keywords: Cyber Security Training Exercise, Cooperation, Behavioral Influence Assessment, System Dynamics, Cognitive Model

INTRODUCTION

Computational models of human behavior can lead to important insights regarding how people interact with each other and with their environments, but validation of these models is often limited. By finding ways to more effectively validate cognitive and behavioral models, the results could become more strongly justified and thus more useful. We studied cooperation between teams in cyber security training exercises by building a model that captures the interactions between them. Real-world exercises provided a useful source of validation data and served to help calibrate the model. Insights learned from the model are intended to help to improve the future versions of the exercise.

Tracer FIRE (Forensic Incident Response Exercise) is a cyber security training program developed by Sandia National Laboratories and Los Alamos National Laboratory. It combines traditional classroom and hands-on training with a competitive game forum. Participants work in teams to solve a series of challenges based on real-world incidents. Teams are sometimes chosen by the participants themselves (which often results in teams segregated by organizational affiliation) and sometimes chosen by organizers based on the expertise of the participants. One planned goal of Tracer FIRE is to improve learning by promoting cooperation during the exercises. The organizers

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2095-4

hope this might improve participants' desire to cooperate in real security situations (which can increase effectiveness and reduce effort required to combat threats) and help to build relationships between participants during the Tracer FIRE exercises. Sandia National Laboratories built a model of team interaction during the Tracer FIRE exercises to provide insight into the drivers and effects of cooperation in this learning environment. The model utilizes the Behavioral Influence Assessment (BIA) framework, a hybrid cognitive-system dynamics structure for simulating systems that involve human behavior and decision making. The theoretical framework of the BIA is based on psychological, social, and economic theories that have been incorporated into a single structure that is both self-consistent and dynamic. Cognitive models are implemented using system dynamics and embedded into an encompassing system dynamics model, which simulates interactions between people, groups, and physical, economic, or other system components.

We collected data from Tracer FIRE exercises to motivate, calibrate, and validate a model of team interaction dynamics. Data included specific decision-making strategies of subject matter experts based on the BIA structure (including cues, perceptions, motivations, intentions, and potential behaviors), environmental data (such as distance between teams and noise level), personality data, and game data (scores, etc.). This information was incorporated into the computational model and used to conduct sensitivity analysis of the information/data and uncertainty quantification of the model output.

This project and the Tracer FIRE BIA model provided insight into how teams made decisions about cooperation and how cooperation might affect performance and learning during the exercise. The model and assessment were used to assess Tracer FIRE challenge designs that would promote cooperation and learning. Simulation exercises proved to be a highly useful source of validation data for this project, and future work is planned to assess how validation data collected from simulations might be used to inform models of real-world cyber security work. Data collected for this model, as well as insights gained from the model, were used to inform a subsequent model of a real-world counterpart of the Tracer FIRE exercises: an information sharing program for cyber defense.

COOPERATION IN CYBER DEFENSE AND THE DATA PROBLEM

Cyber attacks pose a major threat to modern organizations. These attacks can have nefarious aims and serious consequences, including disruption of operations, espionage, identity theft, and attacks on critical infrastructure. Organizations must put substantial resources into protecting themselves and their customers, clients, and others against cyber attacks. However, even with a substantial investment in cyber defense resources, the risk of harm from a cyber attack is significant for many organizations. The effectiveness of cyber defense can likely be enhanced if programs are implemented that allow organizations that face similar cyber threats to share information and resources. The threats faced by different organizations may be similar or even identical (figure 1). Thus, much of the work done by cyber defenders at these organizations may actually be redundant (Hui et al., 2010). By sharing personnel and information regarding effective defense strategies pertaining to cyber attacks, organizations may better protect themselves against cyber threats while maintaining or even reducing the resources dedicated to cyber security.

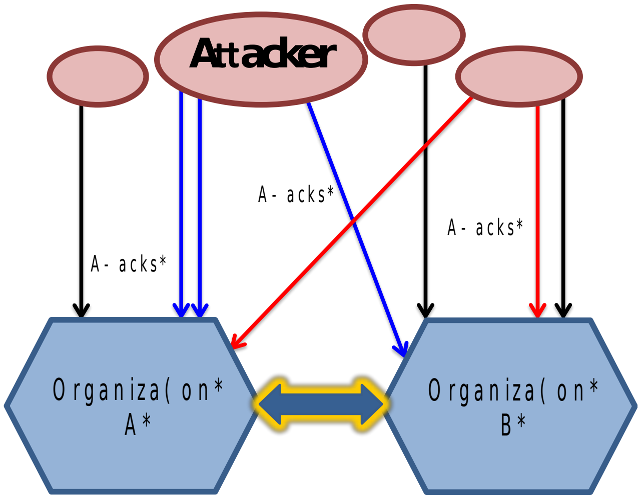

Figure 1: Cooperation can guard against attacks with similar traits or sources

Despite these potential benefits, cooperative cyber defense strategies are not common. Cyber defense teams must balance the potential benefits of cooperation against motivations not to cooperate. For example, if its vulnerabilities are made publicly known, an organization might become more susceptible to cyber attacks and might face damage to its reputation. Trust in cooperating organizations is therefore necessary for successful cooperative cyber security programs. Since organizations that are likely to cooperate with each other are those that face similar threats, they might also be in similar industries and have competitive relationships. Competition for customers, clients, or funding may raise concerns about motive and competitive advantage, making organizations less likely to trust each other. Finally, group inertia is a significant factor to overcome and individual habits may be even more difficult to change than organizational strategy.

This project was designed to begin elucidating how human decision making regarding participation affects an information-sharing program for cyber defense. We created a model that simulates information sharing between multiple organizations. The model focuses on how staff and management in these organizations make decisions about whether and how much to participate in the program. In each organization, decision makers weight the risks and benefits of participation, and their desires to contribute determine the organizations' participation rates. As we built this model, we realized there was a severe lack of information concerning how people might make decisions in these programs. Only a few functional information-sharing programs exist, and the people involved are often wary to share insights on how they do their jobs. Furthermore, it is difficult to collect information on decision making in general (since the mechanisms behind the decisions are often hidden and often not fully understood, even by the decision maker).

To study ways to improve cooperation we used of a training program for cyber defenders called Tracer FIRE. We worked with Tracer FIRE organizers to collect data that could be used in a model of the exercises as well as to inform the information-sharing model. We believe that using controlled exercises as data collection test beds for real-world counterpart systems has immense potential. The exercises were controllable, so that data on behavior determination in different situations could be collected. They also provided a relatively large number of participants who were willing to serve as participants, and who were also willing to discuss their decision making strategies with us at length.

THE TRACER FIRE EXERCISES AND COOPERATION

Sandia National Laboratories and Los Alamos National Laboratory, realizing the increasing threat from cyber attacks, created a training program called Tracer FIRE (Forensic and Incident Response Exercise) to increase the effectiveness of cyber security incident response teams (CSIRTs). Tracer FIRE combines traditional classroom and hands-on training with a competitive game forum. In the classroom portion, students cover incident response topics and are given hands-on training with tools commonly used by CSIRT personnel. In the game portion of the exercise, students form teams and use these tools to solve a series of challenges based on real-world incidents. The challenges cover a variety of cyber defense topics, and the number of points awarded is based on the difficulty of the challenge. The size of the teams varies from 4-10 players. An effort is made to ensure that each team has a balanced skill set and that all teams have roughly the same skill level. Tracer FIRE has been used to train almost 1,000 incident responders from the Department of Energy and other U.S. government agencies, critical infrastructure teams, and academia. In fact, the most recent Tracer FIRE event was held online and had hundreds of participants from over 10 countries around the world. Tracer FIRE also presents an opportunity for human-focused research on cyber security and training. The exercise offers a controlled environment with a variety of challenges and an opportunity for data collection that does not often exist in traditional security environments. A variety of research projects have used Tracer FIRE to study individual and group characteristics in relation to the effectiveness of cyber defense and training.

Tracer FIRE has begun to explore incorporating challenges that encourage cooperation between players. By cooperating with other organizations (sharing information about cyber attacks, effective defense strategies, and personnel with specific expertise), cyber defenders might increase the resources and information available for solving a particular cyber problem and thus better protect their organizations. Researchers have begun to explore the possibility of organizational cooperation in cyber defense (Hui et al., 2010; Sandhu et al., 2010; Luna-Reyes, 2006; Ring & Van de Ven 1994; Oliver, 1990; Luna-Reyes et al., 2008), and the Tracer FIRE team is exploring methods for enhancing cooperation both during and after the exercise. The current design of Tracer FIRE encourages cooperation within teams (points are rewarded by team) and does not prohibit cooperation between teams. Some teams do cooperate with each other to solve challenges, but the point structure, combined with a tendency toward a culture of individualistic work in cyber security (Gates & Whalen 2004), does not always encourage high levels of cooperation.

We collected a large amount of data from the Tracer FIRE exercises to inform, populate, and validate the Tracer FIRE model described below. Our ultimate goal was to inform the information sharing model. We conducted extensive interviews with three Tracer FIRE participants who also work as cyber security professionals. We also conducted shorter, more informal interviews with many other Tracer FIRE participants. We observed multiple rounds of Tracer FIRE and collected data on levels of interaction within and between groups, including personality survey data, and data on environmental condition—such as how close groups were located to each other, the degree of ambient noise, lighting in the room, and the location of shared information. This enabled us to examine two cyber attack (consistent versus uneven attack) scenarios.

THE TRACER FIRE BEHAVIORAL INFLUENCE ASSESSMENT (TF-BIA) MODEL

In order to study the dynamics of cooperation in Tracer FIRE, the Tracer FIRE Behavioral Influence Assessment (TF-BIA) model was created. The model was populated based on interviews with subject matter experts, who were past participants in the Tracer FIRE program and cyber security professionals. It was calibrated using data collected during Tracer FIRE exercises. The model is based on the BIA framework, which was designed to model decision making using well-established psychological, social, and economic theories, all within a system dynamics structure.

Behavioral Influence Assessment (BIA)

Behavioral Influence Assessment (BIA) is a system dynamics-based modeling framework for simulating systems that involve human behavior and decision making. The theoretical framework of the BIA is based on well-established psychological, social, and economic theories that have been incorporated into a single structure (figure 2) that is both self-consistent and dynamic. BIA uses a hybrid, cognitive-system dynamics architecture. Cognitive models are implemented using system dynamics and embedded into an encompassing system dynamics model, which simulates interactions between people, groups, and physical, economic, or other system components.

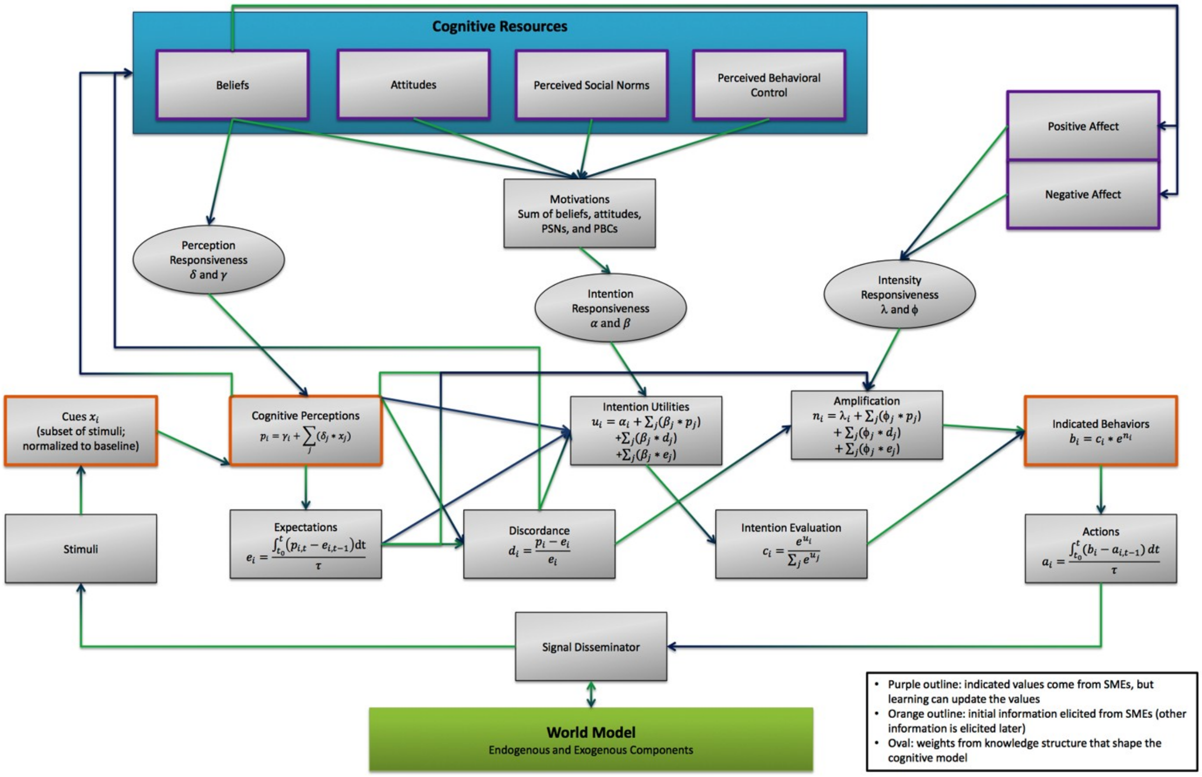
Cognitive Resources
Beliefs
Attitudes
Perceived Social Norms
Perceived Behavioral Control
Positive Affect
Negative Affect
Motivations
Sum of beliefs, attitudes, PSNs, and PBCs
Perception Responsiveness
δ and γ
Intention Responsiveness
α and β
Intensity Responsiveness
λ and ϕ
Cues x_i
(subset of stimuli; normalized to baseline)
Cognitive Perceptions
$p_i = \gamma_i + \sum_j (\delta_j * x_j)$
Intention Utilities
$u_i = \alpha_i + \sum_j(\beta_j * p_j) + \sum_j(\beta_j * d_j) + \sum_j(\beta_j * e_j)$
Amplification
$n_i = \lambda_i + \sum_j(\phi_j * p_j) + \sum_j(\phi_j * d_j) + \sum_j(\phi_j * e_j)$
Indicated Behaviors
$b_i = c_i * e^{n_i}$
Stimuli
Expectations
$e_i = \frac{\int_{t_0}^{t}(p_{i,t} - e_{i,t-1})\mathrm{d}t}{\tau}$
Discordance
$d_i = \frac{p_i - e_i}{e_i}$
Intention Evaluation
$c_i = \frac{e^{u_i}}{\sum_j e^{u_j}}$
Actions
$a_i = \frac{\int_{t_0}^{t}(b_i - a_{i,t-1})\,dt}{\tau}$
Signal Disseminator
World Model
Endogenous and Exogenous Components
• Purple outline: indicated values come from SMEs, but learning can update the values
• Orange outline: initial information elicited from SMEs (other information is elicited later)
• Oval: weights from knowledge structure that shape the cognitive model

Figure 2: Computational structure of the BIA framework

The cognitive portion of the BIA begins with individuals or groups being exposed to cues (stimuli relevant to the decision-maker). These cues are processed to create cognitive perceptions, the decision-maker's assessment of the world or situation. Over time, cognitive perceptions become expectations, which are compared to cognitive perceptions to determine discordance with the current situation. Discordance and cognitive perception affect beliefs, a category of cognitive processes that includes the components of the theory of planned behavior (attitudes, social norms, perceived behavioral control) (Ajzen, 1991) and affect. Intentions are calculated using utility functions. A multinomial logit function (McFadden, 1982) compares intentions to determine realized behaviors, and over time those behaviors become physical realized actions. One of these cognitive models is populated for each individual or group being included in the system. These cognitive models are connected to each other and to a world model sector using system dynamics. The world model sector includes all of the non-cognitive components of the system of interest, including physical systems, economics, etc. Outputs from the world model and the cognitive models act as inputs, or stimuli, for the cognitive model in subsequent time steps. For a broader discussion of BIA see Bernard, Backus, & Bier and Bernard & Bier in this issue of the proceedings).

Tracer FIRE BIA (TF-BIA)

The Tracer FIRE BIA (TF-BIA) model uses the BIA framework to simulate behaviors of participants in Tracer FIRE. The model simulates six teams, each with the same basic cognitive structure (cognitive parameters can vary between teams). Each team determines the amount of effort it spends working individually versus working cooperatively with other teams. Considering the difficulty of the remaining challenges, individual and cooperative progress is calculated. Cooperative progress also takes into account the amount of work required to cooperate with other teams and shared knowledge available through cooperation. Shared knowledge available depends on the amount of knowledge that each team has and the effort that each team puts toward cooperation.

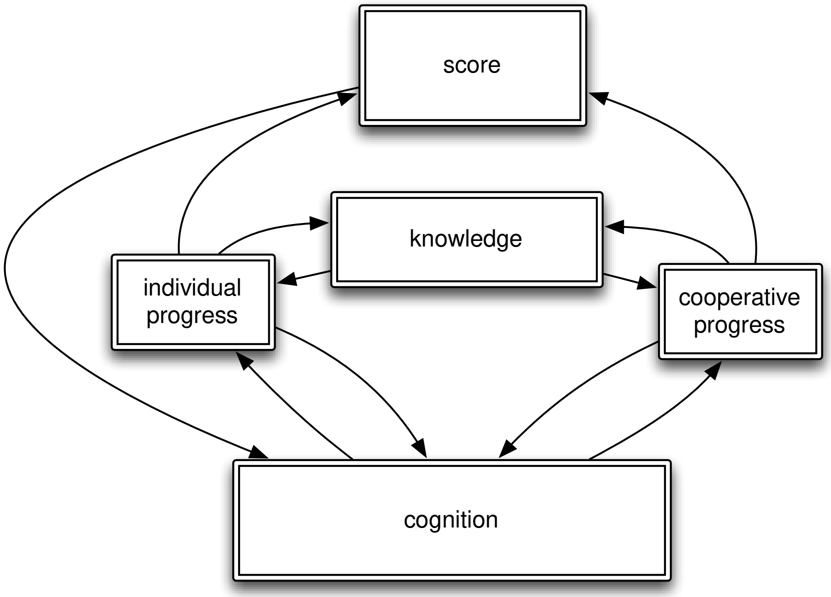

Figure 3: Model structure overview

Individual and cooperative progress for each team is combined to determine the increase in the overall score. As teams solve more challenges, remaining challenges become more difficult. Increase in score and challenge difficulty are used as indicators to determine learning for each team. As knowledge increases, teams become more efficient at solving problems and have more to contribute to cooperative efforts if they choose to do so.

Both behavioral and non-behavioral portions of the model feed into the cognitive models as cues. Interviews with subject matter experts (SMEs) were held to determine how decisions are made during Tracer FIRE. The SMEs were previous participants in the exercise and also work as cyber security professionals. These interviews were used to determine the structure of the decision process (which cues and perceptions are considered, how cues determine perceptions, etc.) and to understand the relative importance of each input for model parameterization. The cues and cognitive perceptions that feed into each potential behavior are shown in table 1.

Table 1: Cues, cognitive perceptions, and potential behaviors

	Potential behaviors →	Work Individually			Work Cooperatively	
	Cognitive Perceptions →	Competitio n	Benefit of indiv. work	Time pressure	Benefit of cooperation	Frustration
	Effect on behavior →	+	+	+	+	+
Cues	Score difference from nearest competitor	-				
	Team rank	+				
	Recent individual progress		+			
	Recent cooperative progress				+	
	Recent total progress					-
	Difficulty of remaining tasks					+
	Time remaining in game			-		

Each team determines how much effort it puts into individual versus cooperative work. Teams tend to increase individual work when they feel time pressure, competition (based on team rank and having competitors close in score), or when individual work has increased the team's score in the recent past. They tend to work cooperatively when they are frustrated (due to lack of progress or high task difficulty), or when cooperation has recently produced benefits. These factors are compared to determine the effort that goes toward each type of work (individual and cooperative), which then affects score and knowledge, as described above.

Select model results

A key goal of Tracer FIRE participants is to win the game (by generating a higher score than any other team), but the primary goal of Tracer FIRE is to increase participants' knowledge about cyber security incident handling. Cooperation allows teams to learn from others, but requires effort and may give competitors an advantage. Teams must decide how much effort to put into cooperation versus individual work, and this decision affects both learning and scores.

There are four adjustable inputs in the TF-BIA model. The first two, initial knowledge (for each team) and baseline cooperation (for each team) are characteristics of the teams but can be altered by the Tracer FIRE designers. In the simulations discussed here, we assume that all teams have the same initial knowledge and baseline cooperation unless otherwise indicated. The other two variables of interest can be directly manipulated by the white cell (the people running Tracer FIRE). The white cell can modify the difficulty of the challenges, which is represented in the model by a maximum task difficulty variable. It can also make it easier or more difficult for teams to cooperate with each other. This might involve changes to communication infrastructure (instant messaging, shared message boards, etc.), locating players in the same room, challenges that encourage cooperation between teams, verbal encouragement to cooperate from the white cell, or other strategies.

The base case simulation is shown in figure 4. In the base case, each team begins with 25% of the knowledge necessary to complete all of the Tracer FIRE challenges. Work required to cooperate is 25% (in other words, only 75% of the effort put into cooperation actually goes toward progress in the exercises). Challenge difficulty is .75 (of a maximum of 1), and each team begins the exercises with a baseline 25% of effort going toward cooperation. The teams end up with about 78% of the maximum score and about 52% of the total knowledge that can be gained from the exercises, doubling their knowledge over the course of the exercise. Cooperative effort starts out at 25% (the baseline), but declines after the beginning of the exercise. Since all the teams have similar, relatively low levels of initial knowledge, not much can be gained from cooperation and teams put more focus into individual work. Competition remains stable in this scenario because the teams' scores are equal. Near the middle of the time horizon, learning and frustration encourage more cooperation. All teams are gaining knowledge, so the potential benefit of cooperation is increasing. The remaining challenges are getting harder (teams tend to solve the easiest challenges first), so frustration is also increasing. At the end of the exercises, time pressure causes teams to focus more on individual work.

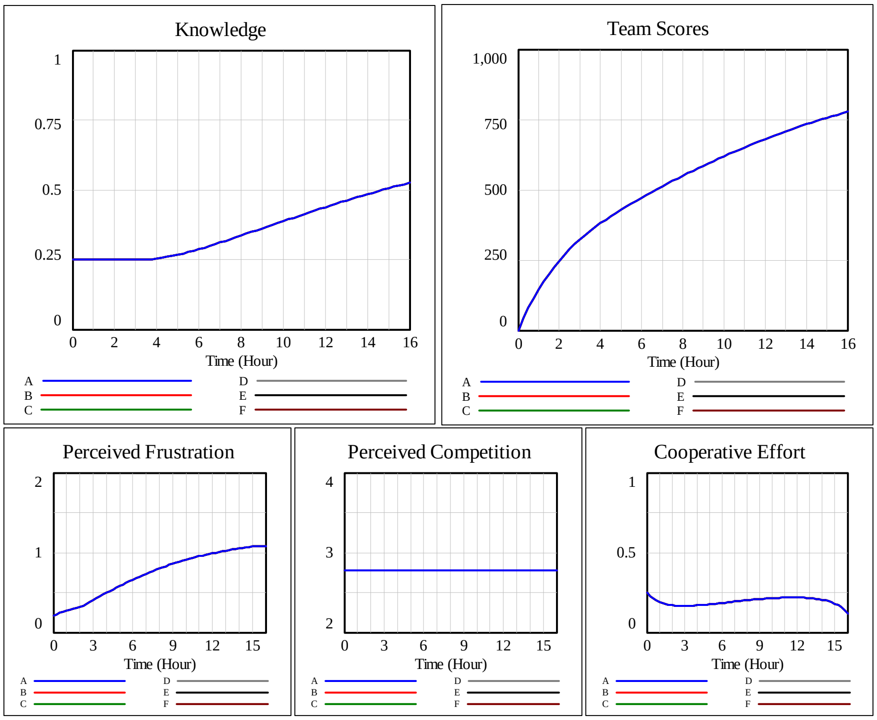

Figure 4: Base case simulation (initial knowledge = 0.25, baseline cooperation = 0.25)

Figures 5 and 6 show scenarios where teams have a higher baseline rate of cooperation (50%) than in the base case (25%). This could represent a situation where teams or participants were chosen specifically for characteristics (personality traits, familiarity with other players, etc.) that encourage cooperation. It could also represent an exercise where teams are encouraged to cooperate before the game starts, or where challenges are designed to encourage cooperation between teams. Both scenarios show that learning increases from the base case. The final knowledge variable for each team nears 66% when baseline cooperation increases to 50% (figure 5), and if barriers to cooperation are removed to make work required to cooperate 5% (rather than 25%), knowledge reaches 70% (figure

6).

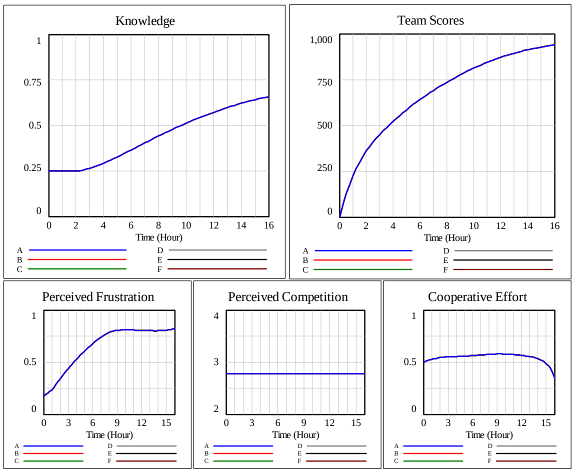

Figure 5: Baseline cooperation = 50%; work required to cooperate = 25%

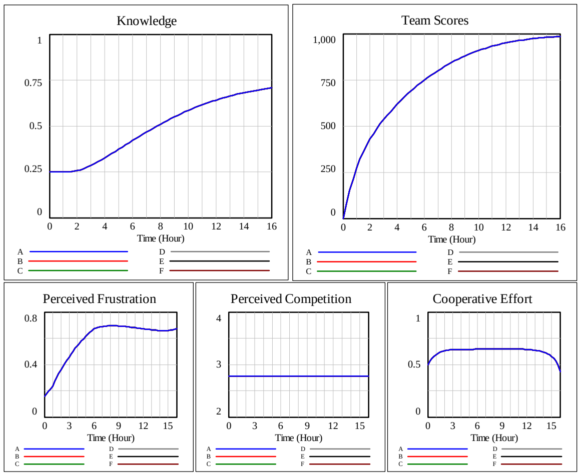

Figure 6: Baseline cooperation = 50%; work required to cooperate = 5%

Learning can be further improved by increasing the difficulty of tasks, as in the scenario shown in figure 7. This scenario is the same as the one shown in figure 5, except that the task difficulty is at its maximum. Participants learn more with higher task difficulty in this scenario, but frustration is also higher. This could cause participants to reduce overall effort levels or to dislike the Tracer FIRE program, discouraging their colleagues from participating in the future. While this model does not consider distraction or future participation in the program, it is a consideration for exercise design and implementation.

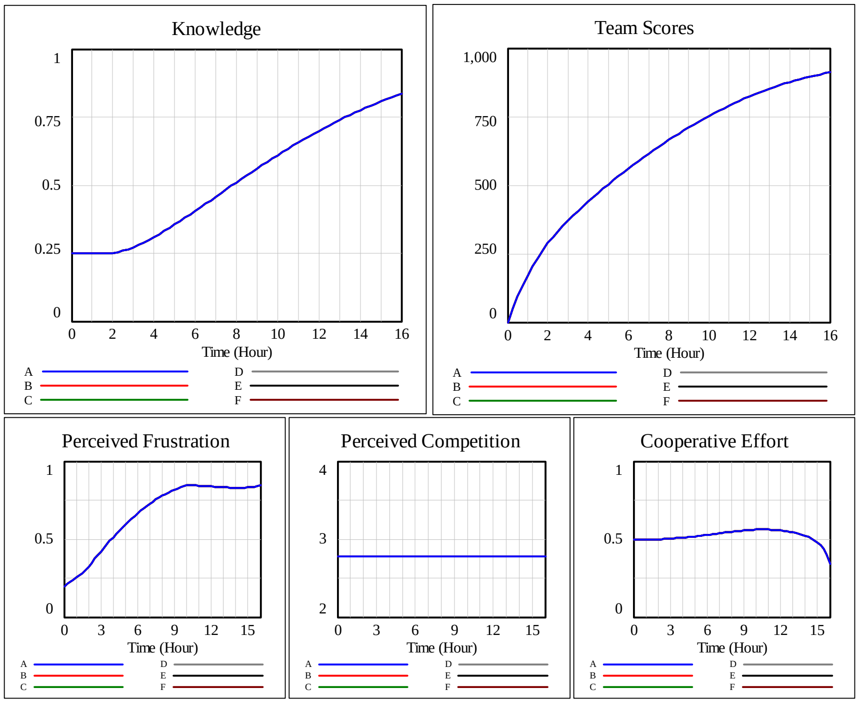

Figure 7: Baseline cooperation = 50%, task difficulty = 1

It is also likely that different teams will have different baseline cooperation levels. Figures 8 shows a scenario in which five teams have baseline cooperation of 25% and one team has a higher level of baseline cooperation (50%). Learning and score both increase a small amount for the team that cooperates more than the others. Figure 9 shows a scenario in which three of the six teams have the higher (50%) baseline level of cooperation. Because more teams are more willing to cooperate, the pool of shared knowledge increases and these teams see an even higher increase in score and knowledge than the others. These scenarios assume that work required to cooperate is the same as in the base case. As barriers to cooperation increase, benefits of cooperation will decrease, at some point (around 50% work required for cooperation in this scenario) creating a negative incentive to cooperate.

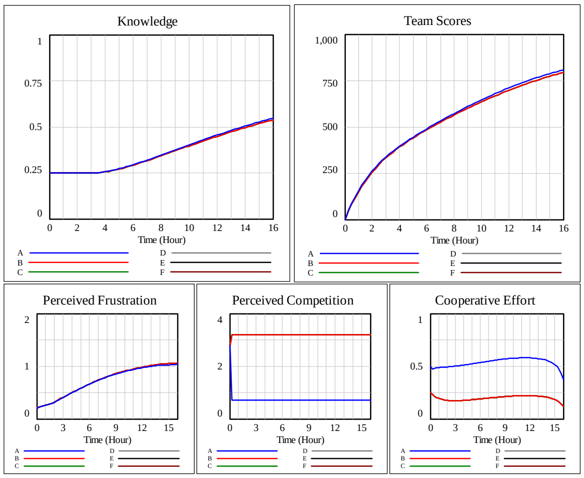

Figure 8: One team with baseline cooperation = 50%

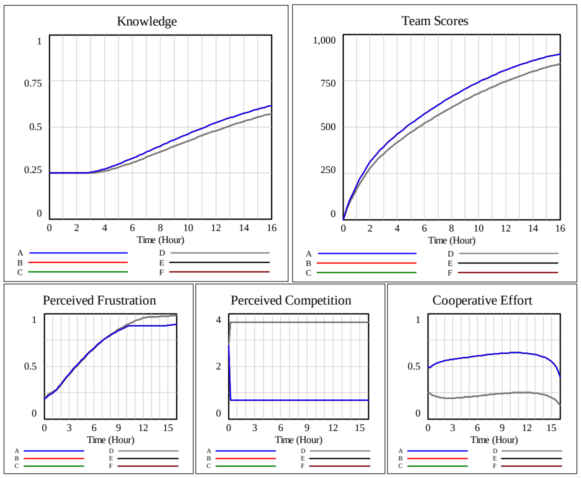

Figure 9: Half of teams with baseline cooperation = 50%

The goal of Tracer FIRE is to increase the participants' knowledge about cyber security incident response. Sensitivity analysis was conducted to indicate which of the four adjustable inputs to this model were most important in determining the teams' average knowledge at the end of the simulation. Partial correlation coefficients are shown in table 2. All of the inputs have high correlation with the knowledge output with high confidence. The maximum task difficulty has the highest (negative) correlation, but the others are also important.

Table 2: Partial correlation coefficients for average knowledge at end of simulation

Variable	Partial correlation coefficient	p-value
Maximum task difficulty	-0.93516	7.8392e-90
Work required to cooperate	-0.92539	4.2709e-84
Average initial knowledge	0.81709	1.5894e-48
Average baseline cooperation	0.75821	4.5148e-38

CONCLUSION: DATA COLLECTION AND VALIDATION

Tracer FIRE presented a great opportunity for human-focused research on cyber security and training. The exercise offers a controlled environment with a variety of challenges and an opportunity for data collection that does not often exist in traditional security environments. A variety of research projects have used Tracer FIRE to study individual and group characteristics in relation to effectiveness of cyber defense and training. These results suggest various strategies that the white cell might try to improve learning during Tracer FIRE. They might make challenges more difficult, remove barriers to cooperation, increase the initial knowledge of participants, or increase participants' baseline levels of cooperation by altering teams based on personality types of participants, composition of teams, familiarity of players with each other, structure of the game, or other strategies. The BIA framework proved useful for modeling cooperative behavior in the Tracer FIRE exercises. It is even more useful in providing a cognitive structure that can be applied both to the exercises and to the real-world information sharing program model. Because the framework includes an explicit cognitive model, we can use the model to understand intermediate phases in participants' decision-making process, such as cognitive perceptions, affect, and motivations. This might be more useful for understanding problems like learning than the decision rule method most common in system dynamics models. The BIA framework shows promise for modeling human behavior, especially in situations where details of cognition may be important.

Using a simulation exercise to study decision making strategies and collect data for models of real-world counterparts proved highly useful for this project. We believe that this strategy can be applied in many other situations. The exercises were alterable, so that we could put decision makers into scenarios that would provide us with the most useful data possible. The large number of participants and relatively casual nature of the situation made it easy for the researchers to discuss decision making strategies with participants. Finally, the large potential for data collection during the exercises (including information on actions taken by participants, scores, and environmental data) provided much more information than would have been readily available directly relating to the information sharing programs we were ultimately interested in.

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.

Bernard, M. L., & Bier, A. B. (2014). Analytical capability to better understand and anticipate extremist shifts within populations in failing states. In *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014.*

Bernard, M. L., Backus, G. A., & Bier, A. B. (2014). Behavioral Influence Assessment (BIA): A multi-scale system to assess dynamic behaviors within groups and societies across time. In *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014.*

Gates, C., Whalen, T. (2004). Profiling the defenders. *Proceedings of the 2004 workshop on new security paradigms (NSPW '04). ACM,* New York, NY, USA, 107-114.

Hui, P., Bruce, J., Fink, G., Gregory, M., Best, D., McGrath, L., Endert, A. (2010). Towards efficient collaboration in cyber security. *International Symposium on Collaborative Technologies and Systems (CTS),* 489–498.

Luna-Reyes, L. F. (2006). Trust and collaboration in interagency information technology projects. *Proceedings of 2006 International Conference of the System Dynamics Society*, Nijmegen, The Netherlands.

Luna-Reyes, L. F., Black, L. J., Cresswell, A. M., Pardo, T. A. (2008). Knowledge sharing and trust in collaborative requirements analysis. *System Dynamics Review, 24*(3), 265-297. doi:10.1002/sdr.404

McFadden, D. (1982). *Qualitative response models.* In Advances in Econometrics, Ed. Werner Hildenbrand, Cambridge University Press, New York.

Oliver, C. (1990). Determinants of interorganizational relationships: Integration and future directions. *Academy of Management Review*, 241–265.

Ring, P.S., Van de Ven, A.H. (1994). Developmental processes of cooperative interorganizational relationships. *Academy of Management Review*, 90–118.

Sandhu, R., Krishnan, R., White, G. B. (2010). Towards secure information sharing models for community cyber security. *6th International Conference on Collaborative Computing: Networking, Applications and Worksharing* (CollaborateCom). (pp. 1–6).

ACKNOWLEDGEMENTS

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. Approved for unlimited release: SAND: 2014-1679C