Title: Modeling as an Aid to Robust Tactical Decision-Making

Topic 6: Modeling and Simulation

Authors: Gary L. Klein[1], Mark Pfaff[2], Jill L.Drury[1], Jennifer Mathieu[1], John James[1], Paula Mahoney[1], and Loretta More[3]

POC: Jill L. Drury
The MITRE Corporation
202 Burlington Road, Bedford, MA 01730-1420 USA
+1-781-271-2034
jldrury@mitre.org

Authors' affiliations:
[1]The MITRE Corporation
[2]Indiana University Indianapolis
[3]The Pennsylvania State University

Abstract

In C^2 situations, decision makers would love having a crystal ball that describes future events not under their control, and how these events would affect each of the courses of action (COAs) being considered. Further, decision makers would appreciate knowing the potential cost of each COA—e.g, via a metric based on the numbers of resources used and the damage, injury, and deaths that may result. While crystal ball technology remains as elusive as ever, Lempert et al. and Chandrasekaran have developed general methods for identifying robust COAs by using simulation models that determine the plausible consequences of each COA under a wide range of possible futures. Because the simulation models must be run many times for each of the (possibly many) COAs, these techniques are computationally intensive, sometimes taking hours, days or weeks. Since tactical commanders need to make decisions in minutes or seconds, we have been manipulating the models underlying two different simulations to determine what computational shortcuts can be made that do not compromise the decision quality of the recommendations. This paper presents results from our analytical and empirical investigations that are facilitating development of a real-time, tactical decision-support system for emergency preparedness and crisis response.

Modeling as an Aid to Robust Tactical Decision-Making

Abstract

In C^2 situations, decision makers would love having a crystal ball that describes future events not under their control, and how these events would affect each of the courses of action (COAs) being considered. Further, decision makers would appreciate knowing the potential cost of each COA—e.g, via a metric based on the numbers of resources used and the damage, injury, and deaths that may result. While crystal ball technology remains as elusive as ever, Lempert et al. and Chandrasekaran have developed general methods for identifying robust COAs by using simulation models that determine the plausible consequences of each COA under a wide range of possible futures. Because the simulation models must be run many times for each of the (possibly many) COAs, these techniques are computationally intensive, sometimes taking hours, days or weeks. Since tactical commanders need to make decisions in minutes or seconds, we have been manipulating the models underlying two different simulations to determine what computational shortcuts can be made that do not compromise the decision quality of the recommendations. This paper presents results from our analytical and empirical investigations that are facilitating development of a real-time, tactical decision-support system for emergency preparedness and crisis response.

1. Introduction

The information needed to attain situation awareness consists of facts about the environment, which Hall et al. (2007) call the *situation space*. Yet, truly aiding decision makers requires providing them with more than situational facts. Decision makers must choose among the options for action that are at their disposal. This additional view of the environment is termed the *decision space* (Hall et al., 2007). Decision makers must be able to compare these options in the decision space and choose among them, given an analysis of the facts of the situation, which maps from the facts to the consequences of each option. For each option there is a distribution of possible consequences. Each distribution is a function of the uncertainty of the elements in the decision situation (how big is the fire) and the uncertainty regarding executing the course of action defined in the decision option (what percent of fire trucks will get to the scene and when).

An optimal plan is one that will return the highest expected return on investment. However, under deep uncertainty (Lempert et. al. 2006), where situation and execution uncertainty are irreducible, optimal strategies lose their prescriptive value if they are sensitive to these uncertainties. That is, selecting an optimal strategy is problematic when there are multiple plausible futures.

Consider a very simple example. Suppose three is the optimal number of fire trucks to send to a medium-sized fire under calm wind conditions, but if conditions get windy, a higher number of trucks would be optimal. If your weather model predicts calm and windy conditions with equal probability, then what will be the optimal number of trucks? One could expend a lot of effort trying to improve the modeling of the weather in order to determine the answer. Furthermore, just how certain are you that the initial reported size of the fire is correct?
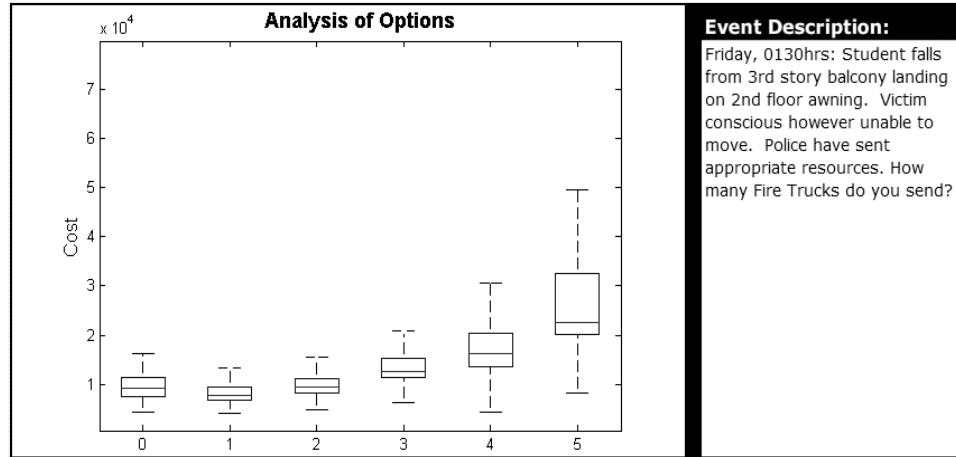
Figure 1: A decision space visualization

Alternatively, Chandresekaran (2005) and Chandresekaran & Goldman (2007) note that for course of action planning under deep uncertainty one can shift from seeking optimality to seeking robustness. In other words, one could look for the most robust line of approach that would likely be successful whether or not it will be windy.

Lempert, et al. (2006) describes a general simulation-based method for identifying robust strategies, a method they call robust decision making (RDM). Using our simple example, one would translate sending different numbers of fire trucks into the parameters of a model. Then for each option, one could explicitly systematically manipulate the other uncertainties of the model (e.g. weather). The model would be executed for each combination of number of trucks sent and set of uncertainties to determine which option performs relatively well across the range of the plausible futures that the model projects. This approach can also identify vulnerabilities of these options, showing under what plausible circumstance each does well or poorly. In turn, this can suggest new options to try to better hedge against these vulnerabilities. Ultimately, this approach enables decision makers to characterize the trade-offs involved in their decision space of different options.

Figure 1 illustrates the results of applying this approach to a fire truck decision in an emergency response situation. Each option for a number of fire trucks to send, which is listed along the horizontal axis of the graph, was analyzed in a simulation model. The hyperspace of futures under a given option is summarized in Figure 1 as a box plot. The box plot is used here merely for illustration, as a common visualization option that typical research subjects can be readily trained to read. Of course, more complex decision spaces will require more domain-specific decision visualization methods. Uncertainty around each value of this *endogenous* variable, such as the actual number of trucks that would arrive in time, was estimated and systematically varied across multiple executions of the model. In addition, other *exogenous* variables that would not be under the control of a course of action, but would likely interact with it (like the wind in our simple example), were also systematically varied across these multiple executions of the model. The result was a hyperspace of combinations of different endogenous and exogenous variable values, which can be considered a hyperspace of plausible future situations. Each of these situations can then be evaluated in terms of how much cost (in this case, the cost of immediate and future damage and injury) is generated by that situation. The result of these evaluations for the hyperspace of futures under a given option can be summarized graphically.

When the cost of each situation is mapped against each course of action, we get a two-dimensional projection that allows us to compare robustness in the users' decision space, as illustrated by the box plots in Figure 1. In this case, across all of the plausible futures, sending one fire truck seems to result in not only a lower median cost (the line inside the box), but also a relatively small range between the cost of the worst cases and the cost of the best cases (the distance between the "whiskers"). In other words, sending one fire truck seems to be the most robust decision option regarding the uncertainties inherent in the situation.

This decision-space construction raises questions regarding the required precision and fidelity of complex simulations that actually are needed to support such robust decision spaces. This is important because modeling an extensive set of possible future conditions can be computationally expensive. In our work we have found it is not unusual to need millions of runs through a model, requiring days or weeks, to compute the consequences of the options for some decisions. However, if we can avoid needless fidelity that does not *substantially* change the decision space, then we can save model development cost and computational time, which in turn will support more tactical decision-making. Certainly, one such substantial change would be to cause a change in the rank order of the options due to changes in the estimated median costs or range of costs. It could also be a change in the distance between options' median costs or ranges. Finally, it could also mean a change in the situations that are source of each cost point.

As part of a research program examining both mathematical and psychological effects, we are performing similar studies on very different models to determine whether we can derive general principles or guidelines for eliminating needless fidelity. Furthermore, we aim to develop a methodology so that others can quickly determine the fidelity boundary conditions and breakpoints for their own models. Once we determine reasonable fidelities for underlying models, we are running human-in-the-loop experiments to determine the effect on decision-makers of providing this information. And while we performed the work in the context of emergency response, we believe the principles will be applicable to other domains in which decision-making is of similar complexity and must take place on similar timescales, such as military command and control.

2. Use Case

Decision-making situations can be described in terms of "VUCAD" characteristics: Volatility, Uncertainty, Complexity, Ambiguity and Delayed feedback (Streufert and Satish, 1997). A fire in a large building is a good example of a difficult decision-making situation because it has challenges pertaining to all of these characteristics. There is:

- volatility regarding the size and location of the fire from one minute to the next,

- uncertainty concerning the presence or absence of hazardous and/or accelerant substances,

- complexity generated by the different alternatives for entering the building and approaching the fire,

- ambiguity regarding the structural soundness of the building at any given moment as the fire attacks its load-bearing elements, and

- delayed feedback to an incident commander located outside the building.

Upon receiving the 911 call indicating a large structure fire, an initial set of resources are sent to the scene. In accordance with the Incident Command System, the highest ranking responder takes responsibility as the Incident Commander (Bush, 2003). In our example use case, Fire Chief Brown becomes Incident Commander (IC) and assesses the situation. Chief Brown considers (Dunn, no year given):

- Is the roof stable? Can I cut a roof vent opening or will it collapse?

- Do I have sufficient resources at the scene or should I call for reinforcements?

- Do I have enough resources at the scene to attack the fire or should I order all firefighters to do search and rescue and let the building burn?

- Should I give up the original burning building and protect adjoining structures with hose lines?

Some other basic questions include (Ennis, 2008): Is there a life hazard? Is the fire spreading?

In this example use case, the fire is already visible from the street through the blown out windows on the second floor. Chef Brown does not have a lot of time to make decisions. He considers how many resources he has at his disposal, the number of possible victims, the potential for property damage, and the likelihood of future events and their potential to draw on his resources. Rather than having to mull over these criteria, he quickly enters into his hand-held PDA the latter three criteria (the PDA already knows how many resources are available back at the station house and at the locations of his mutual aid partners). The resulting display, similar to Figure 1, appears within a few seconds and shows that the most robust approach is to request two additional ladder trucks.

Chief Brown calls for the two ladder trucks and also decides to send a crew inside the building. As a result, Brown's firefighters quickly subdue the fire and keeps it from spreading to the surrounding buildings.

Our work aims to develop a command and control decision aid that is tactical in nature: one that can provide real-time support to decision-makers in high-stakes, time-sensitive situations. We believe that robust decision-making techniques can help us develop these decision support aids.

3. Work Program

The robust decision making technique introduced at the beginning of this paper requires information generated by underlying models. As previously noted, modeling an extensive set of possible future conditions can be computationally expensive. But are all these runs really needed? Must the model be run at a high level of fidelity and precision? If the answers to these questions are "yes," then this RDM approach is untenable for tactical decisions that must be made in seconds or minutes as opposed to days or weeks. One purpose of this investigation was to determine whether the answers to these questions could be "no": that high fidelity and precision is not always needed.

Accordingly, we performed sensitivity analyses to determine points at which the COAs' costs under a lower level of fidelity differed substantially from the costs of the same COAs under a higher level of fidelity. By substantially, we mean that the ordering of the preferred COAs changed from one granularity level to another: which would mean that decision-makers could be led to make the wrong decision.

We started with a model of emergency response called NeoCITIES, described in the next subsection. But to determine whether the results can be generalized to models in other domains, we also examined models of pandemic influenza crisis management (described in section 3.2). Besides modeling different phenomena, the models are of different types (algorithmic/time-stepped, and agent-based).

Once we understood a tactical approach for modeling, we could begin to think about designing a decision aid and testing it with potential users. Section 3.3 describes the human-in-the-loop testing done to date. We end with our conclusions so far and the next steps for this work program.

3.1 NeoCITIES Model Manipulation

A first test case of this research was the model developed for the NeoCITIES scaled-world simulation (Jones et al., 2004). In the words of the NeoCITIES developers: "NeoCITIES is an interactive computer program designed to display information pertaining to events and occurrences in a virtual city space, and then test team decision-making and resource allocation in situations of emergency crisis management" (McNeese et al., 2005, p. 591). Small teams interact with NeoCITIES to assign police, fire/rescue, or hazardous materials assets in response to emerging and dynamic situations. If team members fail to allocate appropriate resources to an emergency event, the magnitude of the event grows over time until a failure occurs, such as the building burning to the ground or the event timing out.

In this work, sensitivity analyses were performed to determine the levels of fidelity at which the options' apparent costs under a lower-fidelity model differed substantially from the costs of the same options using a higher-fidelity model. The high fidelity NeoCities model of the growth of the magnitude of an event is time step-based and incremental: it depends upon the magnitude calculated for the previous time step $t - 1$ and the number of emergency resources (such as fire trucks) applied at that moment. This means that the magnitude of an event at a given time t cannot be obtained without calculating all of the magnitudes at all of the previous times, which is very computationally intensive. Two non-incremental equations were derived from the data generated by the incremental NeoCITIES escalation equation. These lower-fidelity models are computationally simpler. Moreover, being non-incremental, these equations can calculate the magnitude of an event for any time without calculating previous time steps, and therein reduce computational costs.

Although neither of the lower fidelity models were representative of the incremental generating process, they still provided a high fidelity model of the behavior of the data, accounting for 87% and 96% of the variance. Even so, there was a small but significant reduction in discrimination among the decision options with the non-incremental models. This fact illustrates that, even when the behavior of a process can be adequately modeled, there still may be implication for supporting the decision maker. The main effect for the low fidelity model to exaggerate cost, compared to medium or high fidelity, is significant but with a very small effect size ($\eta_p^2 = .001$). The interaction with the courses of action, however, revealed that the two non-incremental models provided less distinct differentiation between courses of action than the "ground-truth" of the original NeoCITIES formula.

The precision with which the situation data was represented in these models was also manipulated. This representation was manipulated along a number of dimensions, the first of which is how accurately the initial magnitude of the event is sampled. Each simulated event's

initial magnitude was selected from a normal distribution for four initial magnitude ranges, one for each dataset: 1 to 2, 2.5 to 3.5, 4 to 5, and 1 to 5. These ranges were partitioned into 3, 8, or 16 values for the low, medium, and high levels of precision, respectively. The remaining aspects of precision are the number of time steps between magnitude measurements (10, 5, and 1, again for low, medium and high) and the number of passes through the simulation (250, 500, or 1000 events).

The manipulation of precision had a significant impact on replicating the decision space. The main effect of precision was that higher levels of precision resulted in less ambiguous decision spaces, that is, where the differences between the consequences of each option were more apparent. In addition, this effect was exacerbated by the uncertainty of the initial estimate of the magnitude of the event. When the estimate was more uncertain, the difference between the highest and lowest precision spaces was greater than when uncertainty was less. That low quality data results in more decision ambiguity is not surprising. However, these results suggest that one way to counter such uncertainty is to engage in a more extensive precise exploration of the distribution of plausible consequences.

This first experiment demonstrated an example where using simpler, lower fidelity, consolidated models of a phenomenon significantly statistically changed the decision space. It is described in more detail in Klein et al. (in press). The next experiment extends these constructs regarding fidelity and precision to a different modeling environment (for pandemic influenza) with the intention of deriving general principles and/or a methodology to determine the boundary conditions of where models can and cannot provide decision-quality data.

3.2 Pandemic Influenza Model Manipulation

A second test case of this research was the high fidelity hybrid disease spread (Beeker, 2009) and process (Mathieu et al., 2009) model developed by MITRE. This hybrid model is designed to run simulations for evaluating planned responses for pandemic influenza. The hybrid model is a discrete-event, agent-based model. The heart of the hybrid model is the calculation of infectivity (i.e., disease spread) and the course of action-related time delays (i.e., process model). The calculation of infectivity is at the person (i.e., agent) level—it dictates the spread of the disease over time given initial disease-related parameters (i.e., strain infectivity, season, population susceptibility, and mortality rate). A key behavior explored via the model is social distancing (i.e., individual compliance)—a voluntary behavior in which people stay home to avoid propagating the disease. The calculation of the effect of process delays is also at the person level, including individuals (1) seeking medical attention (i.e., surge capacity), (2) using antivirals for treatment for the first wave, and (3) obtaining vaccination for the second wave.

The hybrid model drives the distribution of costs for each course of action in the decision space (i.e., level of antivirals, use of social distancing, level of vaccination, and vaccination strategy). To map this situational information into the decision space a way of costing the results of each option was needed. The cost of antivirals (1st wave), vaccines (2nd wave), and productivity loss (both waves) was scaled to the actual population (i.e., 75,000 people). Productivity loss was calculated for both people-sick days and death, with a very high value assigned to death.

Running the simulation also drives the computational cost of each run in terms of computer processing resources needed. Therefore it is the precision of this hybrid model that is manipulated in this study. Two datasets were generated to investigate the effect of precision using this high fidelity model. Each dataset contained multiple predictions of cost for each of the

sixteen possible courses of action for a given pandemic influenza event. This data was generated according to one level of fidelity (i.e., hybrid model) and two levels of precision. Precision in this study has multiple aspects, the first of which is how accurately the parameters and courses of action are sampled. Each was selected from a normal or triangular distribution that represented a 20% range for that individual value. The high precision case utilized the selection from the distribution and the low precision case limited the selection to 3 values. In addition, the high precision case had 800 families (~2170 agents) with 30 replicates, and the low precision case had 400 families (~1085 agents) with 15 replicates.

The model parameters and courses of action were based on Epstein et al. (2008). Future work includes collaborating with these authors to investigate their equation-based model (i.e., lower fidelity) to explore the effect of fidelity and precision on recommended courses of action.

3.2.1 Pandemic Influenza Results

To test how different levels of precision may lead to different cost predictions, a 2 (precision: low, high) x 2 (social distancing: true, false) x 2 (daily vaccination strategy: true, false) x 2 (level of vaccination: 25%, 75%) x 2 (level of anti-virals: 10%, 50%) full-factorial ANOVA was performed. Main effects for three of the four courses of action were found to account for a large amount of the cost for an event: the 75% level of vaccination ($M = 8.89 \times 10^8$, $SE = 1.33 \times 10^7$) led to much lower cost than the 25% level ($M = 9.99 \times 10^8$, $SE = 1.33 \times 10^7$), $F(1,688) = 34.91$, $p < .001$, $\eta_p^2 = .05$; social distancing greatly reduced cost (true: $M = 7.71 \times 10^8$, $SE = 1.33 \times 10^7$; false: $M = 1.12 \times 10^9$, $SE = 1.33 \times 10^7$), $F(1,688) = 340.79$, $p < .001$, $\eta_p^2 = .33$; while daily vaccination increased cost (true: $M = 9.68 \times 10^8$, $SE = 1.33 \times 10^7$; false: $M = 9.21 \times 10^8$, $SE = 1.33 \times 10^7$), $F(1,688) = 6.29$, $p < .05$, $\eta_p^2 = .009$.

An interaction between the level of vaccination and social distancing indicated that the benefits of the higher vaccination level were eliminated when individuals practiced social distancing, $F(1,688) = 13.33$, $p < .02$, $\eta_p^2 = .009$. These results are in Table 1 and Figure 2.

Table 1: Social Distancing X Level of Vaccination Interaction for Cost

Social Distancing	Vaccinating 25%		Vaccinating 75%	
	M	*SE*	*M*	*SE*
False	$1.21 \times 10^9{}_a$	1.87×10^7	$1.03 \times 10^9{}_b$	1.87×10^7
True	$7.92 \times 10^8{}_c$	1.87×10^7	$7.50 \times 10^8{}_c$	1.87×10^7

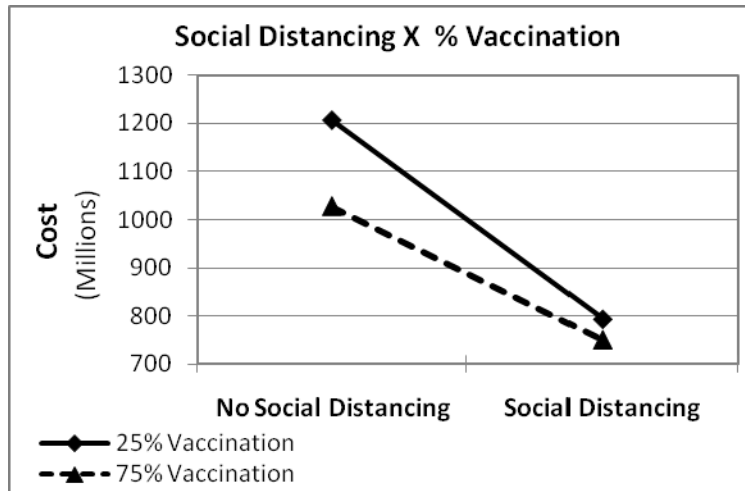Means not sharing a letter differ per Tukey's HSD, α=.050

Figure 2: Social Distancing X Level of Vaccination Interaction for Cost

Precision by itself did not significantly impact cost ($p = .57$), but two interactions with precision were found that did. The first was an interaction between precision and the level of vaccination, shown in Table 2 and Figure 3, revealing that precision moderated the apparent impact of the level of vaccination, $F(1,688) = 4.54$, $p < .05$, $\eta_p^2 = .007$. The benefit of the 75% level is exaggerated in the low-precision condition, as is the increased cost at the 25% level, as compared to the levels observed in the high-precision condition.

Table 2: Precision X Level of Vaccination Interaction for Cost

Precision	Vaccinating 25%		Vaccinating 75%	
	M	*SE*	*M*	*SE*
Low	$1.01 \times 10^9{}_a$	2.16×10^7	$8.63 \times 10^8{}_b$	2.16×10^7
High	$9.83 \times 10^8{}_a$	1.53×10^7	$9.14 \times 10^8{}_b$	1.53×10^7

Means not sharing a letter differ per Tukey's HSD, $\alpha=.050$
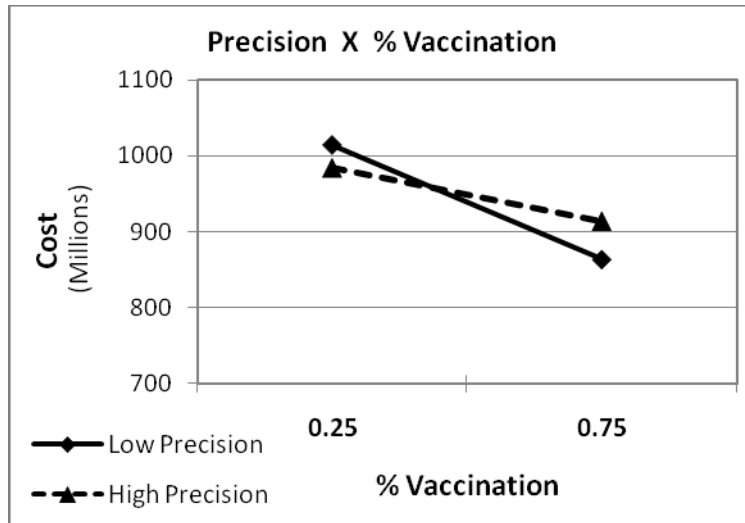
Figure 3: Precision X Level of Vaccination Interaction for Cost

This interaction was moderated further by a three-way interaction between precision, level of vaccination, and social distancing, $F(1,688) = 4.15$, $p < .05$, $\eta_p^2 = .006$. The results are displayed in Table 3 and Figure 4.

Table 3: Precision X Level of Vaccination X Social Distancing Interaction for Cost

| Precision | No Social Distancing | | Social Distancing | |
	Vaccinate 25%	Vaccinate 75%	Vaccinate 25%	Vaccinate 75%
Low	$1.24 \times 10^9{}_a$ (3.06×10^7)	$9.83 \times 10^8{}_b$ (3.06×10^7)	$7.88 \times 10^8{}_c$ (3.06×10^7)	$7.44 \times 10^8{}_c$ (3.06×10^7)
High	$1.17 \times 10^9{}_a$ (2.16×10^7)	$1.07 \times 10^9{}_b$ (2.16×10^7)	$7.97 \times 10^8{}_c$ (2.16×10^7)	$7.56 \times 10^8{}_c$ (2.16×10^7)

Standard error appears in parentheses under each mean. Means not sharing a letter differ per Tukey's HSD, α=.050
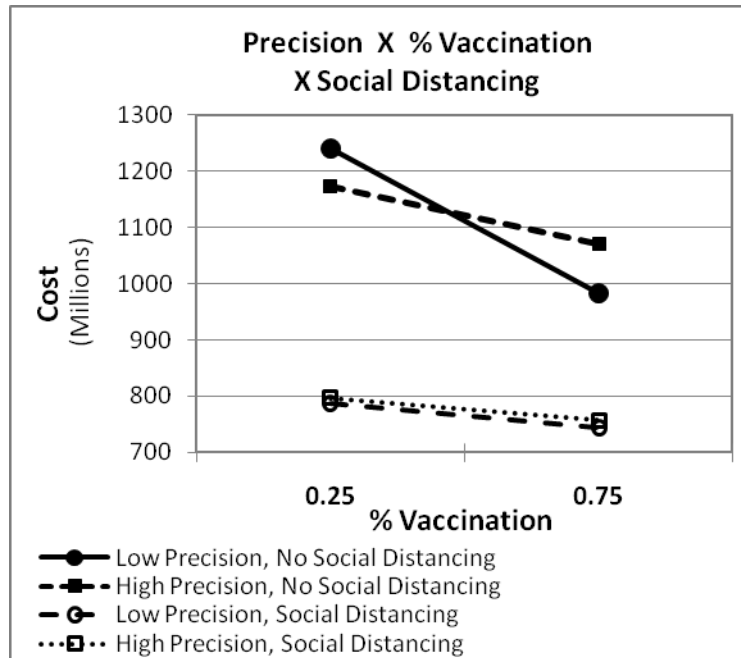
Figure 4: Precision X Level of Vaccination X Social Distancing Interaction for Cost

As Figure 4 indicates, the interaction between precision and level of vaccination only occurs when social distancing does not occur. Otherwise, apart from the clear main effect of social distancing, there is no apparent difference in the cost predictions of either the high- or low-precision levels.

3.2.2 Pandemic Influenza Model Summary

The main effects and interactions found for the courses of action are not surprising, but do serve to validate the functioning of this model. It is the interactions between the manipulations of the model – in this case levels of precision – and these various courses of action that can demonstrate whether such manipulations may lead decision makers to make incorrect decisions.

The first example of this is in the interaction between precision and the level of vaccination. The high-precision level shows a smaller benefit of implementing a higher level of vaccination. Although this may lead to over-expectations of the impact if one was to use the lower precision model, the benefit of the higher level of vaccination is statistically significant regardless of precision used.

Another course of action plays a role in this relationship, which is social distancing. The interaction between social distancing and the level of vaccination, described above, indicates how social distancing effectively negates the need for the higher level of vaccination – the same reduction of cost is achieved either way. Similarly, the interaction between precision and level of vaccination disappears under social distancing conditions: the high- and low-precision levels describe almost identical levels of cost. Once again, the impact of precision did not impact the ordering of the options in the decision space – normatively a decision maker should make the same choice under either model.

3.3 Human-in-the-Loop Testing

Our ultimate goal is to develop decision aids that present potential courses of action available to emergency responders or other command-and-control decision-makers. To determine whether these aids improve decision quality, however, we first needed to develop test scenarios that were challenging in well-understood ways to ensure that we have tested under the full breadth of representative decision-making situations. Thus, we devised a three-step method of developing scenarios, called a Principled Ambiguity Method of scenario design. This method is described in more detail in Drury et al. (2009) but is summarized here as part of providing a complete overview of our inter-related research activities.

The first step is to define the decision space (Hall et al., 2007). As introduced above, the situation space is the factual description of a situation such as surveillance, sensor and alert information; and the decision space is the information needed to make decisions. Put another way, the decision space consists of the results of transforming the raw situation data into something that characterizes the possible courses of action (COAs) in a way that helps decision-makers choose a reasonable alternative from among these COAs. We used the potential costs of each COA to populate the decision space.

The second step is to determine the cost components of each decision's potential consequences based on the principles of Robust Decision Making (Lempert et al., 2003 and Chandrasekaran, 2007). For example, we determined during our investigation in the emergency responder domain relative costs for each COA as a function of the following six items:

$$\text{Cost} = f\{M_i, R, PD_p, PD_f, I_p, I_f\}, \text{ where}$$

M_i Initial magnitude of the incident
R Cost of sending resources, scaled based on the number of resources allocated
PD_p Property damage costs for the current incident
PD_f Any additional property damage costs for future incidents that occur due to the response made for the current incident
I_p Cost of injuries and/or deaths for the current incident
I_f Any additional costs of injuries and/or deaths for future incidents that occur due to the response made for the current incident

The final step is to choose conflicting pairs of cost components (e.g., a small fire, implying low *property damage*, in a densely inhabited area, which implies high *personal injury*). The basic idea behind developing challenging scenarios is to ensure that the scenarios are based on tradeoffs between two conflicting cost parameters: conflicting in the sense that one cost parameter might indicate one COA, and the other parameter might point towards another COA. The mental work needed to determine the "best" COA is more difficult in these ambiguous cases than if all cost parameters point towards a single COA.

In an experiment to validate this approach (Drury et al., 2009), participants made decisions faster in non-ambiguous (control) cases as opposed to cases that included this principled introduction of ambiguity. In fact, not only were the three ambiguous sets significantly different from the control, but post-hoc analysis showed they were all significantly different from each other as well. By pitting decision cost components against each other in a structured fashion, we developed a reproducible approach for generating experimentally viable scenario events. Our

Principled Ambiguity Method of scenario design is also appropriate for use in other domains as long as they can be analyzed in terms of costs of decision alternatives.

Now that we have a human-in-the-loop testbed and a principled method of developing scenarios, we are beginning human-subjects testing using simple visualizations based on box plots prior to developing and testing more sophisticated decision space visualizations.

4. Summary and Future Work

In the two computational modeling experiments that we have conducted to date, we systematically examined the impact of reducing the fidelity and precision of mathematical models, and the precision of agent-based models of the situation space. These reductions of course led to decreases in computation time for lower fidelity/precision models. However, they also resulted in statistically significant changes in the decision space.

Nevertheless, in both experiments these changes were limited to the distance among the options, but not the ordering of the options in the decision space. This suggests that normatively a decision maker should make the same choices under the less computational intense models as under the high fidelity/precision models. Then again, we know people do not always behave normatively.

Future experiments are planned to determine if indeed the impact on the decision space significantly impacts the psychology of the decision makers, either affecting their choices or their confidence in their choices. In addition, we are conducting at least one more computation experiment on the disease spread models to compare the high fidelity agent-based models presented here with more computationally efficient but lower fidelity equation-based models.

5. Acknowledgments

References

Beeker, E. 2009. The MITRE epidemiological model. *System Engineering Process Office (SEPO) Collaborations*, Volume 6 Number 1, p. 9, available at www.mitre.org/work/sepo/pdfs/collaborations_vol6_num1.pdf, last accessed 20 January 2009.

Bush, G. W. (2003). Homeland Security Presidential Directive/HSPD-5: Management of Domestic Incidents. Washington, D.C.: The White House, released 28 February 2003, available at www.whitehouse.gov/news/releases/2003/02/20030228-9.html, last accessed 19 January 2009.

Chandrasekaran, B. (2005). *From Optimal to Robust COAs: Challenges in Providing Integrated Decision Support for Simulation-Based COA Planning,* Laboratory for AI Research, The Ohio State University.

Chandresekaran, B. & Goldman, M. (2007). Exploring Robustness of Plans for Simulation-Based Course of Action Planning. *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making*.

Drury, J. L., Pfaff, M., More, L., and Klein, G. L. (2009). A Principled Method of Scenario Design for Testing Emergency Response Decision-Making. Accepted for publication in the Proceedings of the 2009 International Conference on Information Systems for Crisis Response and Management (ISCRAM 2009).

Dunn, V. (no year given). Life and Death Decision-Making, May Newsletter by Vincent Dunn, Deputy Chief FDYY (ret.). available at http://vincentdunn.com/dunn/newsletters/may/FDNYHP_13.html, last accessed 19 January 2009.

Ennis, R. (2008). Time-Tested. Fire Chief magazine, 1 August 2008. http://firechief.com/leadership/incident-command/time_tested_0808/, last acccessed 19 January 2009.

Epstein, J.M., Goedecke, D.M. Yu, F. Morris, R.J., Wagener, D.K., Bobashev, G.V. (2007). Controlling pandemic flu: The value of international air travel restrictions. *PLoS ONE* 5:e401. http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000401

Hall, D. L., Hellar, B. and McNeese, M. (2007). Rethinking the Data Overload Problem: Closing the Gap between Situation Assessment and Decision Making, *Proc. of the 2007 National Symposium on Sensor and Data Fusion (NSSDF) Military Sensing Symposia (MSS)*, McLean, VA.

Jones, R. E. T., McNeese, M. D., Connors, E. S., Jefferson, Jr., T., and Hall, D. L. (2004). A distributed cognition simulation involving homeland security and defense: the development of NeoCITIES. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA.

Klein, G. L., Pfaff, M. and Drury, J. L. (in press). Supporting a robust decision-space. Accepted for publication in the *Proceedings of the AAAI Spring Symposium on Technosocial Predictive Analytics*, Stanford, CA, March 2009.

Lempert, Robert J., Steven W. Popper and Steven C. Bankes (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis,* Santa Monica, Calif., RAND MR-1626.

Mathieu, J., James, J., Mahoney, P., Fandozzi, J. (2009). Using agent-based approaches to improve disaster response preparedness. *System Engineering Process Office (SEPO) Collaborations*, Volume 6 Number 1, p. 12. Available at www.mitre.org/work/sepo/pdfs/collaborations_vol6_num1.pdf, last accessed 20 January 2009.

McNeese, M. D., Bains, P., Brewer, I., Brown, C. E., Connors, E. S., Jefferson, T., Jones R. E., and Terrell, I. S. (2005). The NeoCITIES Simulation: Understanding the design and methodology used in a team emergency management simulation. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, Santa Monica, CA.

Streufert, S. and Satish, U. (1997). Complexity theory: Predictions based on the confluence of science-wide and behavioral theories. *Journal of Applied Social Psychology*, 27, 2096 – 2116.