Supporting a Robust Decision Space

Gary L. Klein[1], Mark Pfaff[2], and Jill L. Drury[3]

The MITRE Corporation[1, 3] and Indiana University – Indianapolis[2]
[1]7515 Colshire Drive, McLean, VA 22102 USA
[2] IT 469, 535 W. Michigan Street, Indianapolis, IN 46202 USA
[3]202 Burlington Road, Bedford, MA 01730 USA

Abstract

A decision space is defined by the range of options at the decision maker's disposal. For each option there is a distribution of possible consequences. Each distribution is a function of the uncertainty of elements in the decision situation (how big is the fire) and uncertainty regarding executing the course of actions defined in the decision option (what percent of fire trucks will get to the scene and when). To aid decision-makers, we can use computer models to visualize this decision space – explicitly representing the distribution of consequences for each decision option. Because decisions for dynamic domains like emergency response need to be made in seconds or minutes, the underlying (possibly complex) simulation models will need to frequently recalculate the myriad plausible consequences of each possible decision choice. This raises the question of the essential precision and fidelity of such simulations that are needed to support such decision spaces. If we can avoid needless fidelity that does not substantially change the decision space, then we can save development cost and computational time, which in turn will support more tactical decision-making. This work explored the trade space of necessary precision/fidelity of simulation models that feed data to decision-support tools. We performed sensitivity analyses to determine breakpoints where simulations become too imprecise to provide decision-quality data. The eventual goal of this work is to provide general principles or a methodology for determining the boundary conditions of needed precision/fidelity.

Introduction

There is a large body of research literature surrounding *situation awareness*, defined by Endsley (1988) as the "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". Using this definition, 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 therefore 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?

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 enables decision makers to characterize the trade-offs involved in their decision space of different options.

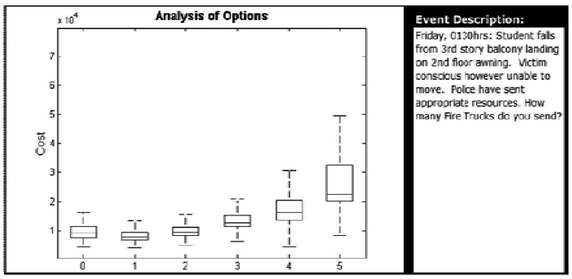

Figure 1: A decision space visualization

Figure 1 illustrates the results of the application of 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 1 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 1 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.

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.

This work constitutes the first phase of a larger investigation. 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. 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.

Background

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.

The NeoCITIES model is designed to run in real time for human-in-the-loop (HIL) simulations. This provides a fast platform to rapidly test the core principles of this modeling research, whereas a high-fidelity version of a disease-spread model we are using in a parallel study can require days or weeks to run. In addition, this work will form the foundation for later HIL research to test the psychological implications of these model manipulations.

The heart of the NeoCITIES model is the magnitude equation, which dictates the growth of emergency events over time given an initial event magnitude score in the range of 1 – 5. It is the core equation in determining the distribution of costs for each option in the decision space. The calculation of this equation also drives the computational cost of each run of NeoCITIES, in terms of computer processing resources needed. Therefore it is the fidelity and precision of this equation that will be manipulated in this study.

The equation 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. The full magnitude equation at time step t is:

$$M_t = (a \times M_{t-1}) + (b \times M_{t-1}^2) - cR$$

where a, b, and c are weights and R is the number of resources applied. We represented the relationship between these factors in a number of computationally simpler ways as described in the methodology section.

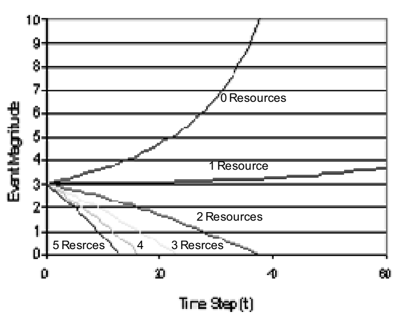

Figure 2: *Relating event-magnitude, time, & resources*

The magnitude equation determined whether assigning a given number of resources R at time step t is effective in resolving the emergency situation. Figure 2 shows a series of six curves that depict the effect over time of applying 0, 1, 2, 3, 4, or 5 resources to a magnitude 3 event at time step t. Sending only 0 or 1 resource results in the event spiraling out of control within 40 time steps in the case of allocating 0 resources or sometime greater than 60 time steps in the case of allocating 1 resource. Sending 2 – 5 resources, results in resolving the emergency within 37 – 13 time steps, with the higher numbers of resources resolving the emergency more quickly, as might be expected.

To map this situational information into the decision space a way of costing the results of each option was needed. NeoCITIES' original scoring equation assigned a combined cost to each resource allocation choice based on the number of expected deaths, injuries, and property damage as well as the cost of sending the designated number of resources. But this formulation did not incorporate the penalty that a decision maker might reasonably expect to pay if he or she assigned too many resources to an event and another event occurred before the first one could be resolved, with the result that the second event suffered more damage because insufficient resources were available nearby. Accordingly, we added to the cost of the current emergency any extra costs that would befall future emergencies due to over-allocating resources to the current emergency.

To determine the distribution of costs for each possible option for a given decision, we ran the model many times each under different conditions that are not under decision makers' control. For example, imagine a fire in which a hot, dry wind fans the flames, versus a sudden downpour that dramatically diminishes the fire. Sending the same number of fire trucks in both situations will result in very different levels of damage; this uncertainty requires a range of costs be calculated for each option rather than a single point value.

Methodology

Development of the models

Two non-incremental equations were developed to model the incremental NeoCITIES escalation equation in a computationally simpler fashion. 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. However, the reduced fidelity of these non-incremental models, as compared to the "ground-truth" of the original NeoCITIES equation, provides one test of the fidelity boundary conditions – where simpler models may lead to recommending different options.

The first equation took a quadratic form:

$$M_t = eM_0 + ft + gRt^2 + h(M_0 t)^2 + i$$

where M_t is the magnitude of the event at time t, M_0 is the initial magnitude of the event at t=0, R is the number of appropriate resources allocated to the event, with e, f, g, and h as weighting constants. This was dubbed the quadratic or "opposing forces" version, in that the force of the escalating event (the *h* term) grows over time against the resisting force of the resources applied to it (the g term). The quadratic form was selected to match the curved trajectories of the original NeoCITIES incremental equation.

The values for the weights were derived via multiple regression of the terms of this equation against the values generated by the incremental NeoCITIES equation. Thus the "ground-truth" data of NeoCITIES was captured in a consolidated model, just as Newton's F=MA captures the data regarding the motion of objects. The NeoCITIES data for the regression was generated over the lifetime of multiple events: five with initial magnitudes ranging from 1 to 5, repeated for each of 6 possible options (allocating between 0 and 5 resources to the event), for a total of 30 runs. Calculations for each run proceeded until the event was successfully resolved (when the magnitude reached zero) or failed by exceeding a magnitude threshold of $M_0 + 1$ or exceeding the time limit of 60 time steps (a typical time limit for events in the NeoCITIES simulation). The results of the regression are in table 1.

	b	*SE b*	*β*
Constant (i)	0.15	0.06	0
e	0.97	0.02	0.84
f	-0.03	1.16e-3	-0.62
g	2.60e-3	5.49e-5	-6.91
h	3.48e-4	7.31e-6	7.35

Note: R^2 = .87. p < .0001 for all β.

Table 1. Regression of quadratic formula weighting constants

A second regression was performed upon a linear form of the above equation:

$$M_t = jM_0 + kt + mRt + ntM_0{}^2 + i$$

where *j*, *k*, *m*, and *n* again are weighting constants. The other variables remained as above. The linear form was selected as the simplest model of the original NeoCITIES incremental equation. The results of this regression are in table 2.

	b	*SE b*	*β*
Constant (i)	0.18	0.04	0
j	0.95	0.01	0.82
k	-6.02e-3	5.31e-4	-0.12
m	-0.08	8.43e-4	-2.52
n	0.01	1.25e-4	2.48

Note: R^2 = .96. p < .0001 for all β.

Table 2. Regression of linear formula weighting constants

Testing the Effects of Fidelity and Precision

Four datasets were generated to compare the three models: original incremental, non-incremental quadratic and the non-incremental linear. Each dataset contained multiple predictions of cost for each of the six possible courses of action (COA) for a given event. This data was generated according to a 3x3 factorial design with three levels of fidelity (the three equations described in the previous sections) and three levels of precision. For fidelity, the original NeoCITIES magnitude equation was assigned to the *high* condition, the quadratic version to the *medium* condition, and the linear version to the *low* condition. Precision in this study has multiple aspects, 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).

Results

One-way ANOVAs were performed for each data set evaluating the distances among the median cost for each option; this quantitative analysis is analogous to a human visual inspection for differences. For example, the way in which the box-plots would be analyzed by a user of the decision aid illustrated in Figure 1. This quantitative analysis was done for all nine combinations of fidelity and precision,.

To test how different combinations of fidelity and precision may lead to different cost predictions, a 3 (fidelity) x 3 (precision) x 6 (course of action) full-factorial ANOVA was performed, controlling for data set. As expected, a main effect for the course of action was found to account for a large amount of the cost for an event, $F(5,20943) = 506.85$, $p < .001$, $\eta_p^2 = .11$. Sending more resources to an event reduced its cost, but this effect reversed somewhat for excessive over-allocation of resources. A main effect was also found for fidelity ($F(2,20943) = 13.38$, $p < .001$, $\eta_p^2 = .001$) with low fidelity having a higher mean cost (*M* = 53694.58, *SE* = 267.19) than medium (*M* = 52335.50, *SE* = 266.84) or high (*M* = 51801.30, *SE* = 266.24; medium and high means were not significantly different per Tukey's HSD, α=.050). An interaction was found between fidelity and course of action, shown in Figure 3 ($F(10,20943) = 8.38$, $p < .001$, $\eta_p^2 = .004$).

The high-fidelity model shows greater differentiation between the mean costs associated with each course of action than the medium or low fidelity models. This was confirmed by analyzing the variances associated with each combination of fidelity and precision using a 3 (fidelity) x 3 (precision) ANOVA, controlling for the data set. A main effect for fidelity on variance was highly significant ($F(2,24) = 10.48$, $p < .001$, $\eta_p^2 = .47$) with the mean variance for the high fidelity condition

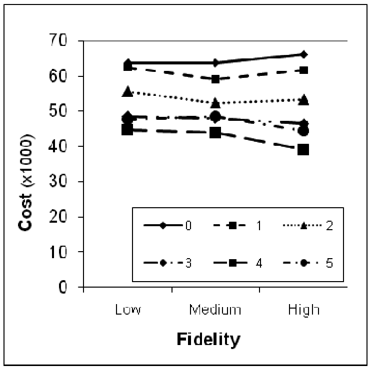

Figure 3. Interaction between Fidelity and COA

significantly higher (M = 4.97e+8, SE = 2.11e+7) than medium (M = 3.64e+8, SE = 2.11e+7) or low (M = 4.02e+8, SE = 2.11e+7; medium and low means were not significantly different per Tukey's HSD, α=.050).

As a measure of between-model ambiguity, the F-ratios for each ANOVA were then compared using a 3 (fidelity) x 3 (precision) ANOVA, again controlling for the data set (these F-ratios were log-transformed for a normal distribution). These results are graphically depicted in Figure 4.

This revealed a highly significant main effect for precision, $F(2,17) = 180.29$, $p < .001$, $\eta_p^2 = .99$. The mean F-ratio for the highest precision level was largest (M = 4.83 (125.71), SE = 0.04), followed by the medium level of precision, (M = 4.21 (67.44), SE = 0.04), and the lowest level (M = 3.64 (38.00), SE = 0.04)[1]. A main effect for fidelity was also significant ($F(2,17) = 44.78$, $p < .001$, $\eta_p^2 = .88$) with the mean for the high fidelity condition significantly higher (M = 4.52 (92.10), SE = 0.04) than medium (M = 4.13 (62.17), SE = 0.04) or low (M = 4.03 (56.26), SE = 0.04; medium and low means not significantly different per Tukey's HSD, α=.050). This indicates that there is a statistically significant increase in ambiguity for using either of the two alternative equations. There was no interaction between fidelity and precision.

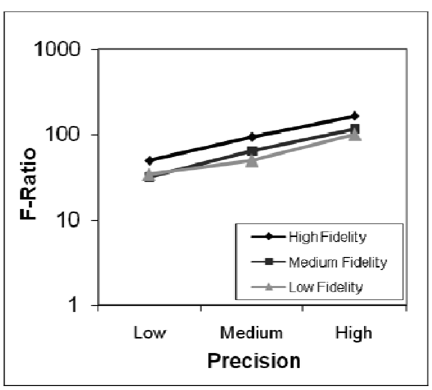

Figure 4. Results of 3x3 ANOVA for precision and fidelity

[1] Back-transformed values appear in parentheses after the mean.

An interaction was found between precision and data set, $F(6,17) = 37.43$, $p < .001$, $\eta_p^2 = .95$. Table 3 presents the results of this analysis that is graphically illustrated in Figure 5. This effect revealed a trend toward greater increases in ambiguity (lower F-ratios) for lower-precision models receiving more uncertain input data. Comparison of the 1 – 5 data set vs. the 2.5 – 3.5 data set is the most meaningful in the present study, as the 1 – 2 and 4 – 5 sets have floor and ceiling effects, respectively.

	1 – 5 Dataset		2.5 – 3.5 Dataset	
Precision	*M*	*SE*	*M*	*SE*
2	3.84 (46.36)[a]	0.08	5.69 (295.05)[d]	0.08
1	2.92 (18.62)[b]	0.08	4.98 (145.01)[e]	0.08
0	1.71 (5.50)[c]	0.08	4.38 (79.51)[f]	0.08

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

Table 3. Precision X Data Set interaction for F-Ratio

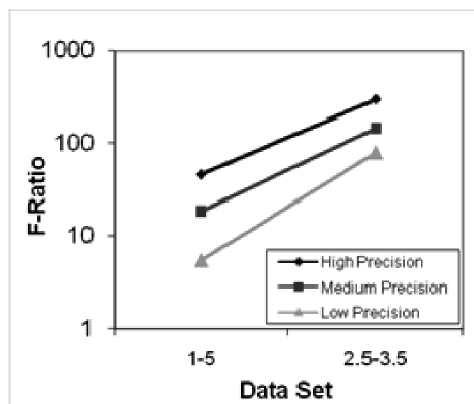

Figure 5. Precision x Data Set interaction for F-Ratio

Discussion and Conclusions

The Effects of Fidelity

We had experimentally defined fidelity in terms of how well the non-incremental formula matched the incremental data-generating formula. The latter being non-linear, we defined the quadratic formula as higher-fidelity. However, intriguingly, it was the linear non-incremental formula that accounted for more of the variance in the behavior of the data that was generated. Ultimately, even though neither of these non-incremental formulae 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 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 regret, compared to medium or high fidelity, is significant but with a very small effect size ($\eta_p^2 = .001$). The interaction with the course of action, however, reveals that the two non-incremental models provide less distinct differentiation between courses of action than the "ground-truth" of the original NeoCITIES formula. Further human-in-the-loop experiments will need to be pursued to determine if these differences are psychologically significant. This could have implications for the utility of using statistical models of social behavior instead of psychological models of social processes.

The Effects of Precision

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.

The results indicate that this experiment's manipulation of precision was too coarse to clearly establish where the boundary lies between precise enough and not enough. Follow up experiments should explore the regions between the values use here.

Conclusions

This work was a first step at addressing the question of the essential precision and fidelity of models that are needed to support decision spaces. We demonstrated an example where using simpler lower fidelity consolidated models of a phenomenon significantly statistically changed the decision space. Whether or not this translates into a substantial psychological change will have implications for whether supporting robust decision making with such simplified models can safely save computational time.

This paper breaks new ground toward an eventual goal of providing general principles and/or a methodology to determine the boundary conditions of where models can and cannot provide decision-quality data.

Acknowledgments

This work was supported by The MITRE Corporation Innovation Program projects 19MSR062-AA and 45MSR019-AA. Thanks to Loretta More, David Hall and Jacob Graham of The Pennsylvania State University (PSU) for acting as sounding boards for ideas discussed in this project; and to the many developers of the NeoCITIES simulation at PSU.

References

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.*

Endsley, M. R., 1988. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meetin*g, Santa Monica, CA, Human Factors Society.

Hall, D. L., Hellar, B. and McNeese, M., 2007. Rethinking the Data Overload Problem: Closing the Gap between Situation Assessment and Decision Making. In *Proeedings 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.

Lempert, R.J., Groves, D.G., Popper, S.W. & Bankes, S.C., 2006. A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios, *Management Science*, Vol 52, No. 4, April, pp. 514-528.

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.

Tukey, J. W., 1977. *Exploratory Data Analysis.* Reading, Mass: Addison-Wesley.