A Principled Method of Scenario Design for Testing Emergency Response Decision-Making

Jill L. Drury
The MITRE Corporation, Bedford, MA
jldrury@mitre.org

Mark Pfaff
Indiana University Indianapolis, IN
mpfaff@iupui.edu

Loretta More
Pennsylvania State University, State College, PA
lmore@ist.psu

Gary L. Klein
The MITRE Corporation, McLean, VA
gklein@mitre.org

ABSTRACT

We are investigating decision aids that present potential courses of action available to emergency responders. To determine whether these aids improve decision quality, however, we first developed test scenarios that were challenging in well-understood ways to ensure testing under the full breadth of representative decision-making situations. We devised a three-step method of developing scenarios: define the decision space, determine the cost components of each decision's potential consequences based on the principles of Robust Decision Making, then 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)*. In a validation of this approach, experiment participants made decisions faster in non-ambiguous cases versus cases that included this principled introduction of ambiguity. Our Principled Ambiguity Method of scenario design is also appropriate for other domains as long as they can be analyzed in terms of costs of decision alternatives.

Keywords

Scenario design, research methods, experiment design, decision support, emergency operations decision-making, robust decision-making, RDM, Principled Ambiguity Method.

THE PROBLEM

Consider a fire chief who must decide how many assets to deploy to a single fire. It is important to allocate sufficient assets to quickly bring the situation under control, but not dedicate unnecessary assets such that responses to potential future incidents are needlessly impaired[1].

Through long experience, emergency responders normally apply heuristics effectively for allocating resources. For example, small fires almost always require a small number of firefighting assets, and large fires normally require more trucks and fire retardant materials. But resource allocation decisions are not always this straightforward. If a small fire is burning in a field well away from the nearest structures or people, the chief might make a very different decision than if the same size fire occurs in an uninhabited jewelry shop or in a crowded classroom building. The reason for treating these situations differently is because the potential cost, in both money and lives, for each of the three situations is quite diverse—even though they all involve the same size fire. Emergency responders consider potential costs when making decisions, so decision-making test scenarios need to reflect this fact.

Emergency responders also pay attention to clues that indicate that there is a high likelihood of further incidents occurring prior to when resources can be released from the current incident. There is a tension between using many resources to rectify the current situation quickly, versus using fewer resources to combat the emergency more slowly but save resources to respond to the next emergency.

Our research goal is to design effective decision aids for emergency responders. To determine whether our aids result in high-quality decisions, however, we need to be able to test them under a representative sample of challenging decisions. This sample needs to be challenging in well-understood ways, such that the set of test scenarios causes experiment participants to handle all of the major types of cognitive challenges for emergency

[1] It usually takes longer for "mutual aid" resources from other jurisdictions to arrive if insufficient resources remain at the primary response location.

response—whatever these are. This paper describes our search for the major cognitive challenges in emergency response, the method for constructing scenarios that encompass these challenges, and an experiment validating our new method.

RELATED LITERATURE

Many researchers have studied decision-making, although fewer have concentrated on decisions in emergency response. Streufert (2005), in particular, provides an excellent survey of research in decision-making under emergency conditions. Streufert, however, does not describe decision-making situations beyond the "VUCAD" characteristics: volatility, uncertainty, complexity, ambiguity and delayed feedback (Streufert and Satish, 1997). Indeed, the majority of the emergency response decision-making literature defines emergency situations and their responses at a very generic level, such as the description by Malakis and Kontagiannis (2008) of cognitive strategies in air traffic control emergencies: recognition, managing uncertainty, planning, anticipation, managing workload, coordination, information exchange, error management, and workload distribution management.

Following a very different approach, the National Oceanic and Atmospheric Administration developed a "Community Vulnerability Assessment (CVA)" technique (NOAA, 2008) that enables civic organizations to determine the degree to which they are responsible for special populations and environmental areas that need a disproportionate amount of attention in a disaster situation. For example, there are sections of the CVA that pertain to identifying "high need" neighborhoods, the community's largest employers, and critical natural resources. While the CVA is not a decision-support tool per se, it has as an underlying assumption the fact that communities should identify in advance the likely geographical areas or populations that would be at particular risk for damage or injury. In other words, the CVA asks communities to identify the people and places in which the costs of an emergency could be higher than the norm.

In line with the idea of understanding the costs of an emergency is the concept of Robust Decision Making (RDM; Lempert et al., 2003 and Chandrasekaran, 2007). An optimal course of action (COA) is always extremely situation-dependent and is often sensitive to conditions beyond decision-makers' control. Under the deep uncertainty (Lempert et al. 2003) that is characteristic of crisis management situations, decision-makers would be better served by making *robust* decisions that are less sensitive to inaccuracies in situation descriptions. Alberts and Hayes (2003) defined robustness as the ability to maintain effectiveness across a range of tasks, situations, and conditions; they consider it one of the six dimensions of agility. Lempert et al. (2003) and Chandrasekaran (2007) both describe 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.

RDM models output a range of costs for each potential COA, after taking into account the range of circumstances that could make each COA either inexpensive or expensive. The advantage of giving decision-makers a range of costs instead of a single cost for a decision alternative is that the decision-maker can determine which decisions are more likely to stay within the range of acceptable costs, regardless of external factors not under the decision-maker's control. In other words, such an approach could help decision-makers choose the alternative that is most reasonable under the widest range of circumstances: one definition of the most robust solution.

Our cost function is, in fact, a multi-attributed utility (MAU) function (Chatfield et al., 1978; Keeney and Raiffa, 1993). Holloway notes that, "A major difficulty in making decisions under uncertainty is that 'good' decisions can have 'poor' outcomes and vice versa" (Holloway, 1979, p. 11). The problem with previous MAU representations is that the distributions of these good and poor outcomes and their values are collapsed into a single probability-weighted average for each option. Decision-makers find it difficult to accept and interpret meaningfully this kind of MAU representation. This difficulty is one reason that MAU has been so underutilized in the field by decision-makers under uncertainty (Klein, 1982).

In contrast, the RDM approach maintains additional distribution information and presents it visually to the user. RDM not only shows the median MAU of each option, but also shows the distribution of MAUs that occur in multiple plausible futures. In implementations currently under development we even will be able to allow the user to explore these distributions interactively.

Moreover, RDM can help us understand the cognitive challenges for emergency responders. Decision-making becomes cognitively difficult under conditions of uncertain information or when choices have different but seemingly balanced sets of positive and negative characteristics. RDM can make the pros and cons of decisions more apparent, as we will explain below. It can also help us construct challenging scenarios by elucidating how to play off pairs of scenario attributes, such as a *small* fire in a *highly populated* building.

THE PRINCIPLED AMBIGUITY METHOD OF SCENARIO DESIGN

Our scenario design method takes three steps: define the decision space, determe the decision space cost components, and choose conflicting pairs of cost components.

Step 1: Define the Decision Space

There inevitably exists a gap between the description of a situation and the information needed to make decisions. Hall et al. (2007) call this the Situation Space-Decision Space gap. Decision-makers must make sense of the raw information in the situation space, such as that which is obtained from surveillance, sensors and alerts. In the emergency management example, the situation space information might take the form of a map-based display with symbols indicating locations of fires, motor vehicle accidents, and emergency vehicles. Decision-makers must then mentally transform this information into a decision space containing a set of alternative COAs and their plausible consequences. The decision space consists of the results of transforming the raw situation data into something that characterizes the possible COAs in such a way as to help decision-makers choose a reasonable alternative from among these COAs.

RDM accomplishes a mapping from the situation space to the decision space by employing a simulation model to evaluate each decision alternative against all possible futures and determining a cost range for each alternative. RDM takes into account not only the decision choices, but also incorporates the effects of the variables not under the control of the decision-maker under all plausible circumstances, from best to worst.

In the case of the fire chief example, consider the small fire in the area well away from structures or people. If a warm wind might whip the flames in the direction of the local chemical plant, the chief might make a very different decision than if a drenching downpour were to occur. The chief might want to visualize the potential for weather extremes and their effects on allocating two versus three ladder trucks to the scene. A graphical display based on weather simulation could make it easy to compare the results, using a relative cost metric, of what will happen when allocating one versus two ladder trucks under the best and worst conditions.

In our initial explorations of a decision aid, we wished to make the possible range of costs of each decision more apparent to the decision-makers, so that they can choose an alternative that has an acceptable cost under virtually all potential circumstances. To show the cost ranges, we needed a visual aid that would allow us to show at least the minimum, maximum, and median costs. As a starting point that would enable us to test the concept of RDM for emergency response, we employed a simplified box-plot visualization with no outlier data points (Tukey, 1977) as pictured in Figure 1. (Note that the point of the investigation was to evaluate the use of RDM, not the use of box-plots per se.) In the figure, each box-plot shows the lowest cost for the best possible future for that particular decision alternative as a lower whisker, the 25th percentile cost as the bottom of the box, the median cost as a red line within the box, the 75th percentile cost as the top of the box, and the highest cost as the top whisker. Thus, each individual box-plot shows a range of costs that depict the effects of the variables beyond the decision-makers' control, such as whether high winds whip flames or drenching rain quenches the fire, assuming the same decision has been made in each case.

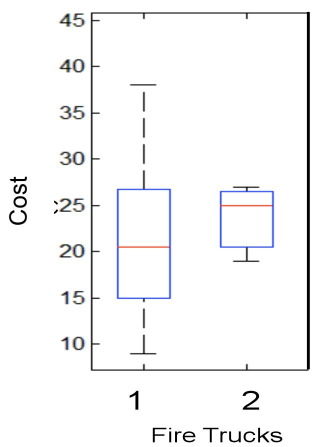

Figure 1 (left). Box-plot visualization showing the relative costs of sending one fire truck versus two assuming a range of possible futures.

Step 2: Determine the Decision Space Cost Components

For our investigation in the emergency responder domain, we determined cost metrics based on a relative scoring algorithm called the *regret equation*. This equation is used both in creating the decision-space visualizations, as depicted in Figure 1, and in understanding how testing scenarios should be developed. In the emergency response case, the costs calculated for each COA are a function of the following six items:

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

We used a model that is algorithmic-based and time-stepped. The size of the incident's initial magnitude, as mitigated by the number of resources allocated to handle the incident, dictates whether it dies out within a reasonable amount of time, takes so long to die out that its response is considered to be a failure, or escalates out of control (also counted as a failure, obviously). The cost of property damage is based on the initial magnitude, the resources applied, and the value of the property that is affected. The costs associated with injuries or deaths are also dependent on the initial magnitude, resources applied, and number of victims affected. Note that some of the costs that will be borne in the future are associated with the current incident if the current incident's response means that there are insufficient resources for the future incident; and there is more damage, injury, or deaths because of the consequent delayed response.

Step 3: Choose Pairs of Conflicting Cost Components

Cost-function conflicts. We hypothesized that the tradeoffs between the cost components in the regret equation would be challenging for emergency responders to weigh, and thus should figure into a test of a decision support aid. We have already alluded to some of these tradeoffs above. Using a lot of resources now versus saving them for future incidents represents a tradeoff between high PD_p and a possible high PD_f and/or between high I_p and a possible high I_f. The basic idea behind developing challenging scenarios is to ensure that the scenarios are based on tradeoffs between two conflicting cost parameters; e.g., 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 conflicted cases than if all cost parameters point towards a single COA.

Table 1. Example Challenging Scenario Components Based on Cost-function Conflicts

No.	Value/Parameter 1	Value/Parameter 2	Example Scenario Summary
1	Low M_i	High PD_p	A small fire has occurred in a jewelry shop
2	Low M_i	High I_p	A small fire has occurred in a crowded classroom
3	High M_i	Low PD_p	A large brushfire is burning well away from structures
4	High M_i	Low I_p	A large fire is burning in an uninhabited warehouse
5	High PD_p and High I_p	Possible High PD_f	A fire is burning in a crowded jewelry shop, but the city has been alerted that terrorists are threatening to bomb a (closed) art gallery
6	High PD_p and High I_p	Possible High I_f	A fire is burning in a crowded jewelry shop, but elsewhere in the city there is a packed stadium and home-town fans may riot if their team loses the game

Table 1 contains no case with High PD_p and/or High I_p versus Low PD_f and/or Low I_f. This omission is due to the fact that this particular scenario is not necessarily challenging. Decision-makers' jobs would be much easier if they knew for a certainty that nothing else would happen that would need their resources before the current incident was resolved, because then it would be a smart choice to apply all available assets to the current incident, regardless of its size.

While Table 1 contains "Low vs. High" (or "High vs. Low") tradeoffs in the first four cases, it is also possible to construct scenarios based on "Medium vs. High" or "Low vs. Medium" parameters using the same approach.

Robustness conflicts. Some of the scenarios we constructed that had both parameters set to "High" (or "Medium") introduced a different kind of conflict. These situations may, as a result of how the model takes into account conditions beyond decision-makers' control, show the cost of the very best case for a particular COA as the lowest but the worst case associated with the same COA may be much worse than that of the other COAs. This "robustness conflict" situation is illustrated in Figure 1.

Figure 1 shows an interesting dilemma in the positions of the whiskers for the two box-plots. Sending two fire trucks would almost certainly result in a higher cost than sending one fire truck under the best circumstances, thus making it less attractive on a best-case basis. However, sending two fire trucks would have much less probability of catastrophic results than sending one fire truck, thus making it the preferred choice on a worst-case basis.

A similar type of robustness conflict is introduced by a box-plot with the lowest 25th percentile case but the highest 75th percentile case (the lower and upper bounds of the box-plot minus its whiskers, respectively). Finally, there may also be robustness conflicts introduced by positions of median cases, shown in the box-plots as lines bisecting each box. Using the example of Figure 1, sending two fire trucks would have a higher median cost than in the case of sending one fire truck, but two fire trucks would still involve much less expense in the worst case.

Because they often seemed to have the characteristic of robustness conflicts, our heuristic for developing these cases was to design a scenario that had the potential for a large amount of damage or many injuries. We ran the model for a number of these scenarios and chose the ones whose box-plots showed evidence of this type of conflict. It is likely that this type of ambiguity is sensitive to whatever algorithm or assumptions are used by the model generating the results, and so cases of this type will need to be developed in an iterative, trial-and-error fashion.

VALIDATING THE PRINCIPLED AMBIGUITY METHOD OF SCENARIO DESIGN

Since we postulated that conflicts among the cost components (which we call ambiguous decisions) would give rise to challenges in mentally trading off one versus another, we defined the following hypothesis:

H = Participants will make non-ambiguous decisions more quickly than ambiguous decisions

Accordingly, we instrumented the test environment to capture decision times automatically.

We recruited 9 female and 12 male participants from a non-profit corporation. Two participants were 30 years old or less, 4 were between 31 and 40, 7 were between 41 and 50, and 8 were between 51 and 60 years old. While we designed the experiment to not require emergency response domain experience, four of our participants had some experience in this domain.

The portion of the experiment used to validate the scenario design method employed a within subjects design. Each participant was asked to make decisions in four different types of scenarios:

A. *Unambiguous*. There was clearly one COA that was better than the others in all dimensions. This set of cases functioned as the control condition.

B. *Ambiguous: magnitude (M_i) versus current cost (PD_p and/or I_p)*. These scenarios were of types 1 - 4 in Table 1.

C. *Ambiguous: current costs (PD_p and/or I_p) versus future costs (PD_f and/or I_f)*. These scenarios were of types 5 and 6 from Table 1.

D. *Ambiguous: best case costs (at the 0-percentile cost and 25-percentile levels) versus worst case costs (at the 75-percentile cost and 100-percentile cost levels) versus median costs*. These scenarios reflected the robustness conflicts discussed in the previous section.

Experiment Conduct

Upon arriving, participants were asked to read a paper copy of a one-page introduction to the experiment, which included Institutional Review Board information. They were given a paper copy of a training manual to read and keep as a reference during the experiment. Next, they were given five training scenarios on the computer interface to become familiar with it. After the training, they completed 40 scenarios (10 of each type described above) during which participants were asked to allocate police or fire/rescue resources. Presentation order of the scenarios was randomized.

Each scenario contained a short textual situation space description of the emergency that included information that suggested the likely cost of the incident (e.g., a fire at a jewelry shop) and the likelihood of another incident occurring soon. Some participants also saw a box-plot diagram of the decision space comparing the COAs. (While box-plots were created using the model, some were adjusted to accentuate ambiguity.) Each scenario was completely independent; what happened in one scenario did not affect another scenario and the number of resources available was reset to the maximum at each new scenario.

After reading the provided information, the participants made a decision regarding the number of resources to send (0 to 5). Immediately after each decision, participants were asked to rate their confidence in that decision on a 7-point semantic differential scale (1 = "low" and 7 = "high").

After completing all of the scenarios, participants answered survey questions. We used three different survey instruments to determine where participants fell on the spectra of risk taking versus risk aversion (Blaise and Weber, 2006), visual versus verbal information processing (Childers et al., 1985), and vivid versus non-vivid imaging (Sheehan, 1967).

Participants were allowed to use work time for participating in the experiment but were given no further compensation.

Experiment Results

Data was analyzed using a within-subjects ANOVA comparing decision speeds[2] between the four sets of scenario events. Decision speed was significantly faster in the Set "A" cases (the unambiguous cases) versus each of the other types of cases, $F(3,816) = 86.97$, $p < .001$, $R^2 = .47$, supporting our main hypothesis. 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.

A second within-subjects ANOVA compared decision confidence for each of the four sets of events. All three ambiguous event sets showed significantly lower confidence than the unambiguous control set, $F(3,816) = 18.97$, $p < .001$, $R^2 = .47$. The results can be seen in Table 2 below. No significant effects were found for covariance of any of the measures of personality traits or prior experience with the scenario text or text-plus-box-plot conditions.

Table 2. Results from Validation Experiment

Set	Set Description	Decision Speed Results	Confidence Results
A	Unambiguous (control)	$M = 9.52$ (13.65 seconds), $SE = .08$	$M = 5.40^a$, $SE = .20$
B	Magnitude vs. current costs	$M = 9.81$ (18.26 seconds), $SE = .08$	$M = 4.91^b$, $SE = .20$
C	Current vs. future costs	$M = 10.05$ (23.17 seconds), $SE = .08$	$M = 4.86^b$, $SE = .20$
D	Best vs. worst vs. median costs	$M = 10.23$ (27.62 seconds), $SE = .08$	$M = 4.75^b$, $SE = .20$
All decision speed means significantly different per Tukey's HSD at α=.05. Confidence means not sharing a letter differ at α=.05.			

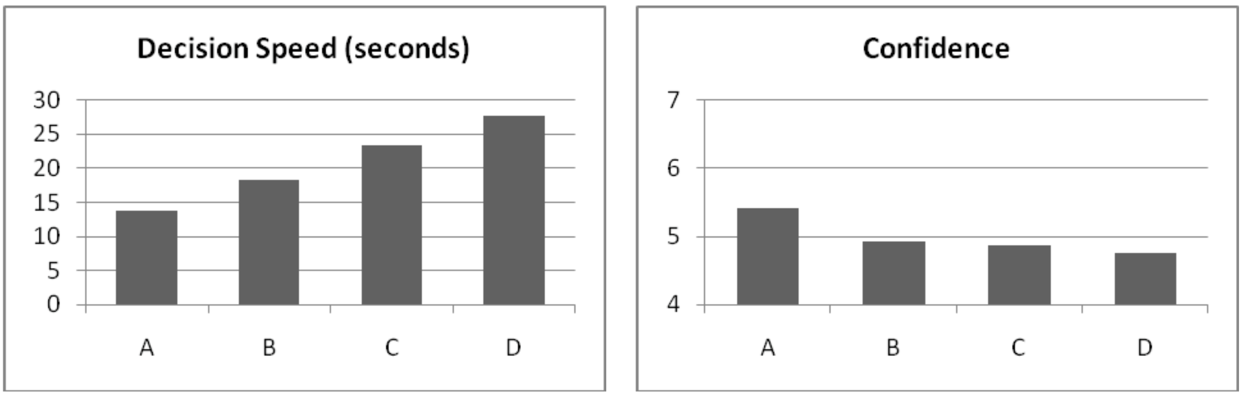

Figure 2. Decision Speed in Seconds and Participants' Assessment of Confidence in Their Decisions
Confidence was rated on a scale of 1 to 7, and the legend for each set of decisions is included in Table 2.

[2] Decision speed was log-transformed for a normal distribution. Back-transformed means appear in parentheses after the mean.

These results demonstrated that scenario events employing these specific ambiguity principles successfully induced the mental tradeoff behaviors we expected, which not only extended decision-making times as hypothesized, but also diminished the level of confidence in those decisions.

CONCLUSION

Our objective in this study was to determine a methodology that scenario designers could use in testing the efficacy of decision aids under a representative sample of challenging decisions. The results in the preceding section demonstrate that by pitting decision cost components against each other in a structured fashion, we have developed a reproducible approach for generating experimentally viable scenario events. The benefit of this methodology lies in how it can administer well-understood doses of ambiguity (that is, the source(s) of the ambiguity are clearly defined, and the amount of ambiguity differs significantly between event types). Armed with this approach, scenario designers can produce libraries of events of known challenge levels, against which experimenters can measure the potential improvements in decision speed and confidence afforded by decision aid systems. This experiment validated three specific event types in our methodology for designing ambiguous scenario events. Future work can explore additional variations and permutations of these tradeoffs.

We are currently developing an analogous set of decision-making scenarios using a different type of model (agent based) using a different domain that is nevertheless still related to emergency response (pandemic flu response). One goal of this ongoing work is to confirm the generalizability of the methods described in this paper. The results so far, however, indicate that our Principled Ambiguity Method of scenario design is promising for use in other domains that can be analyzed in terms of costs of decision alternatives.

ACKNOWLEDGMENTS

This work was supported by The MITRE Corporation Innovation Program projects 19MSR062-AA and 45MSR019-AA. We would like to thank Robert Hooper of PSU for help with scenario development. Thanks, also, are due to David Hall and Jacob Graham of PSU for acting as sounding boards for ideas discussed in this project and for information systems support to conduct the experiment.

REFERENCES

1. Alberts, D.S. and Hayes, R.E. (2003). *Power to the Edge*. Washington, DC: Command Control Research Program.
2. Blais, A., & Weber, E. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*(1), 33-47.
3. 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.
4. Chatfield, D. C., Klein, G. L., Copeland, M. G., Gidcumb, C. F., & Schafer, C., Multiattribute utility theory and conjoint measurement techniques for air systems evaluations. Technical Report, Pacific Missile Test Center, N6126-78-1998, 1978.
5. Childers, T. L., Houston, M. J., & Heckler, S. E. (1985). Measurement of individual differences in visual versus verbal information processing. *The Journal of Consumer Research*, *12*(2), 125-134.
6. Keeney, R.L. and Raiffa, H. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. New York, New York: Cambridge University Press, 1993
7. Klein, G. L. (1982). Heuristic influence on decision processes. PhD. Thesis, Department of Psychology, Texas Tech University.
8. 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.
9. Holloway, C.A. (1979). Decision Making Under Uncertainty: models and choices. Englewood cliffs, New Jersey: Prentice Hall.
10. Lempert, R. J., Popper, S. J. and Bankes, S. C. (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis. Santa Monica, Calif., RAND MR-1626.
11. Malakis, S. and Kontagiannis, T. (2008). Cognitive strategies in emergency and abnormal situations training: implications for resilience in air traffic control. In *Proceedings of the 3rd Symposium on Resilience Engineering*, Juan Les Pines, France.

12. National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center (2007). Common Vulnerability Assessment tutorial. http://www.csc.noaa.gov/products/nchaz/htm/methov.htm, updated 24 July 2007, accessed 4 January 2009.
13. Sheehan, P. (1967). A shortened form of Betts' questionnaire upon mental imagery. *Journal of Clinical Psychology*, *23*(3), 386-389.
14. Streufert, S. (2005). Emergency decision making and metacomplexity. In *Proceedings of the 2nd Annual ISCRAM Conference*, Brussels, BE, April 2005.
15. 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.
16. Tukey, J. W. (1977). *Exploratory Data Analysis.* Reading, Mass: Addison-Wesley.