# "Badge-Primed" Decision Making

**Jill L. Drury, Amanda Anganes, Heather Byrne, Maria C. Casipe, Roger Dejean, Simone Hill, Tristan Lewis, Jesse Lucas, and Eric McCann**
University of Massachusetts Lowell
{jdrury, acortem, hbyrne}@cs.uml.edu, {Maria_Casipe, Roger_Dejean}@student.uml.edu, shill@cs.uml.edu, Tristan_Lewis@student.uml.edu, {jtlucas, emccann}@cs.uml.edu

## ABSTRACT

We have been investigating new decision support methods for emergency responders. Most recently, we have added to our decision support prototype the concept of "badges": symbols that cue decision makers to the top-ranked option(s) that are the recommended alternatives for a particular decision. This paper provides the rationale for badges, a description of the initial implementation, results from our first experiment with badges, and a discussion of the next steps. As a report on work-in-progress, the primary contribution of this paper is the description of the concept of badges and its proposed use for emergency response decision making.

### Keywords

Decision making, option awareness, icons, badges, visualization

## INTRODUCTION

Recognition-Primed Decision Making (RPD) is generally thought to be especially applicable to emergency response. According to Klein (1999), emergency responders employ RPD by observing the facts of a situation, matching them to a mental template developed through experience, selecting a decision option that was acceptable in the past, and mentally simulating the option under current conditions. When situations are very complex, however, simulations may be difficult to run inside decision makers' heads. Thus we have been developing decision support tools to provide cognition aids to emergency responders. We have hypothesized that providing symbols (or "badges") to cue decision makers to top-ranked options in a visualization of decision alternatives can support better performance, and might be called "badge-primed decision making."

To begin constructing visualizations of decision options, we started with Bankes' (1993) Robust Decision Making and Chandrasekaran and Goldman's (2007) exploratory modeling techniques. These techniques take advantage of the fast processing capabilities that are now available to run simulation forecasting models of different sets of plausible assumptions to determine a range of outcomes for each decision option. The modeler does not have to decide in advance which assumptions are most likely to describe the situation, because all reasonable assumptions can be tried in multiple runs of a simulation model. (See Klein et al. 2009 for an analysis that shows high fidelity is not always needed, thus aiding the speed of the simulation models.)

Each of the results of the simulation model runs is scored according to a multi-attributed utility function (Keeney and Raiffa 1993) that is tailored to the decision domain. Plotting these scores graphically means that the emergency responder can see the range of outcomes for one option versus another. The goal is to reduce the decision maker's cognitive load when identifying and evaluating a large number of contingencies by offloading that work to the computer. Thus our approach can enable emergency responders to determine quickly an option in which results are acceptable over the widest range of plausible futures (Pfaff et al. 2010a).

The results of the simulation model runs for each option can be summarized graphically by a box-plot (Tukey 1977). Box-plots for options in an example structure fire are shown in Figure 1. (Note that this example is deliberately simpler than the complex decisions that we eventually plan to support.) The top and bottom "whiskers" of the box-plot indicate the maximum and minimum cost cases, respectively. The top and bottom sides of each box depict the cost of the 75th and 25th percentile cases for their respective options. The line bisecting the box indicates the median cost case. Figure 1 indicates that the most robust option is to send one fire truck from Station A: its box-plot has the lowest cost for three of the five box-plot components (the minimum, 25th percentile, and median). Information that enables decision makers to evaluate the relative desirability of one decision option versus another helps to build "option awareness" (Drury et al. 2009).

In our previous experiments, we found that supporting option awareness by providing the box-plot display in addition to a textual description resulted in 43% correct decisions, versus 31% when participants received only the textual description of the situation (Pfaff et al. 2010a). While this difference was statistically significant, 43% did not seem high enough for real-world operations. Thus we added the ability to change the weights of the box-plot components describing each option depending upon the decision makers' judgment and context.

We defined four weighting strategies. "Equal" consists of the original approach, which applied equal emphasis to all five box-plot components (as implied by a "best 3 out of 5" rule). An "emphasize maximum" weighting strategy puts most weight on having the best maximum value (i.e., the lowest cost maximum), followed by the next highest weight on the 75th percentile cost, etc., with the lowest weight being given to the minimum cost case's value. We also defined an "emphasize minimum" weighting strategy that puts the most weight on having the best minimum value and least weight on the maximum value. Alternatively, a decision maker may wish to minimize costs, which would likely occur with the lowest median value. Thus we defined a "normal" weighting strategy that placed most weight on the median and least weight on the maximum and minimum values.

An empirical evaluation of the visualizations enhanced with the ability to specify different weighting strategies resulted in a significantly higher level of correctness: 62% (Pfaff et al. 2010b), versus 43% when no alternative weighting strategies were provided (Pfaff et al. 2010a). Now we are modifying the visualizations yet again in an effort to further aid emergency responders in choosing the most appropriate option. This paper presents the new design and our empirical results to date.

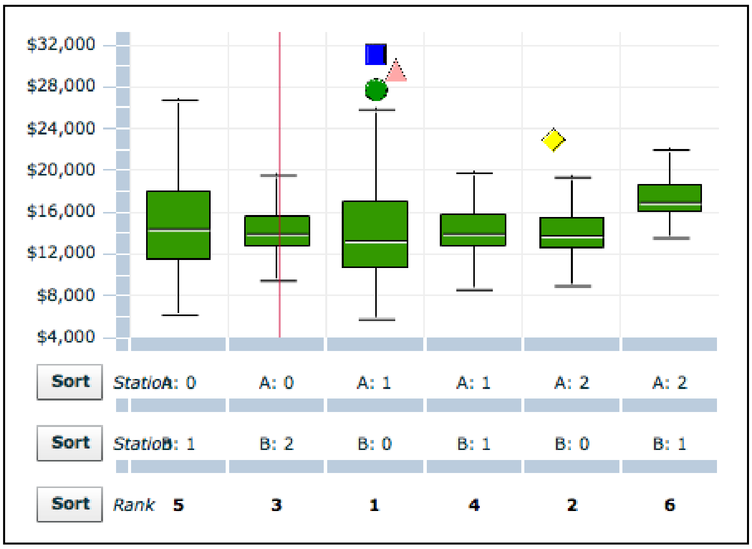


**Figure 1. Box-plots show the results of running multiple cases for each option (sending x fire trucks from Station A and y fire trucks from Station B). Badge symbols indicate the top-ranked option for each strategy. In this example they show that sending 1 fire truck from Station A is the top-ranked option for three of the four weighting strategies.**

## ADDING BADGES TO THE VISUALIZATION

Each weighting strategy has an associated symbol that is unique in both shape and color; that is, symbols are double encoded to ensure those with color vision deficiencies can discriminate among the symbols. The symbol set consists of a yellow diamond, pink triangle, blue square, and green circle. We call the symbols "badges" because this term evokes the positive traits of competence and authority, for example via analogies to the badges worn by firefighters to indicate their role or the "merit badges" earned by youth in scouting programs.

An individual option can rank first in zero, any, or all of the four weighting strategies, meaning there will be between zero and four badges associated with each option. The badges appear directly above the box-plot for the option that ranks first for the weighting strategy indicated by the badge, as can be seen in Figure 1. A legend is located directly below the box-plot ranks (not shown in Figure 1). Badges are arranged in a diamond pattern above the appropriate option, with the same badge always appearing in the same position of the pattern.

It is possible for an option to be top-ranked according to one weighting strategy but be ranked 2 or 3 in another strategy. For example, in Figure 1 the ranks are shown for the "normal" weighting strategy at the bottom of the figure. Sending 2 fire trucks from Station A is ranked #2 under the "normal" strategy but is ranked first under the "emphasize maximum" strategy as shown by the yellow diamond symbol above its box-plot.

If an option is the top-ranked choice in three or four weighting strategies, this would seem to be an important cue for decision makers. In fact, participants in the pre-badge experiments often explored all or several of the four weighting strategies as a way to obtain multiple "opinions" on which option most often had the best rank

(Pfaff et al., 2010b). In other words, experiment participants flipped back and forth between weighting strategies, activating a radio button to indicate a strategy and watching as the ranking labels updated to reflect the strategies' potentially different ordering of the options. This meant that participants had to remember which option was top-ranked for each strategy from one presentation of box-plot ranks to the next.

By making decision makers remember which option is top ranked in one strategy before changing to another strategy, we were causing them unnecessary cognitive load. Indeed, user interface experts recommend that knowledge should be available for people to view "in the world" rather than requiring users to store and retrieve knowledge "in the head" (Norman 1988). Using this principle, the display should allow decision makers to see which option(s) are top-ranked for all strategies at once. We reasoned that if decision makers could see which options are top-ranked for the various weighting strategies in a single glance, without having to toggle the box-plot display among the different weighting strategies, they could more quickly and more confidently make more correct decisions. Thus we modified the interface to include symbols that draw attention to the top-ranked options, and replicated an earlier experiment to determine whether these badges improved performance.

## RELATED WORK

Our decision support tool can be thought of as a visualization-enhanced recommender system; that is, the tool provides a recommendation for which option to choose. We noted two studies in particular that used symbols or icons as part of their recommendation visualization: Campbell (2004) and Gupta and Pu (2003). Campbell investigates the use of icons as ways to draw attention, or point towards a recommended action that automobile drivers should take to maintain safe operation. They concluded that there are no standards and that proposed symbol designs should be tested to ensure their effectiveness. Gupta and Pu (2003) explored how to visualize reputation (the popularity of an item) and recommendations. They tested two different design approaches to depict recommendations used by Citeseer (a scientific literature search engine) and found that the approach that used fewer graphical elements was preferred. Our badges were designed to have a minimum of graphical elements.

We also looked at the literature addressing the design of icons in user interfaces beyond recommender systems (e.g., Christ 1984, Cleveland and McGill 1995, Conati and Maclaren 2008, Mackinlay 1986, Nowell et al 2002). Researchers have evaluated icons based on color, shape, size, and position attributes. In general, each study produced different conclusions, leading us to believe that icons should be tailored to specific applications.

## EVALUATION METHODOLOGY

An earlier experiment (Pfaff et al. 2010b) explored participants' performance when given a decision aid that included the ability to view options' ranks under different weighting strategies. In the current experiment, we used the exact same procedure and decision aid except we added badges to the visualizations, as described above. We compared performance between the group in the 2010 experiment, which we call the "control condition," and the group in the current (2011) experiment, which we call the "badge condition." Thus we employed a between-subjects experimental design, with the independent variable being display presentation (no badges versus badges) and the dependent variables being decision speed, correctness, participants' confidence, and the number of times participants changed weighting strategies.

We defined four hypotheses, as follows:

H1: Participants in the badge condition will choose the correct answer (defined as an option that is top-ranked in at least one weighting strategy) more often than participants in the control condition.

H2: Participants in the control condition (that is, without the extra support offered by badges) will have increased decision-making time compared to participants in the badge condition.

H3: Participants with badges will have more confidence in their responses than the participants without badges.

H4: Participants with badges will change weighting strategies fewer times than participants without badges.

Participants in the control condition were recruited from undergraduate information technology courses at a major northeastern university (N = 27, ages ranging from 18 to 25) and two locations of a not-for-profit corporation (N = 14, ages ranging from 26 to 66). Thirty-seven participants were male and four were female. Ten of forty-one participants had prior familiarity reading box-plots and eight of the forty-one had prior emergency management experience. Data for two participants were eliminated due to technical problems at one of the testing sites.

Participants in the badge condition were recruited from graduate students in a different major northeastern university (N = 15, ages ranging from 18 to 52). Twelve of the participants were male and three were female.

Seven out of the fifteen had prior familiarity reading box-plots and one of the fifteen had prior emergency management experience. Data from one participant was lost due to technical difficulties. Four individual data points were lost for four separate participants.

In both conditions, we used the same script to ensure that all participants received the same information. After signing an Informed Consent form, we provided participants with hardcopies of training materials to go through at their own pace. After this phase of training was complete, we provided another set of hardcopy training materials to guide participants through hands-on use of the decision aid to complete the first of the training scenarios. After finishing the rest of the training scenarios, participants completed twenty scenarios in which they made decisions regarding the correct number of fire vehicles or police squad cars to send to each incident. Finally, participants completed an on-line questionnaire.

## RESULTS AND DISCUSSION

H1 (badge condition participants will make more correct decisions) was not supported. There was a significant difference between the performance for those who received badges, who chose a top-ranked answer 52% of the time, versus 66% for the control group, $p < .00003$.

H2 (control condition participants will take longer to decide) was not supported. On average, participants in the control group needed slightly more time in which to make decisions (33734 ms) than those in the badge group (31731 ms), but the difference was not significant, $p < 0.179$.

H3 (badge condition participants will have more confidence) was supported. Using a Likert scale from 0 to 7 where 0 is not at all confident and 7 is completely confident, the badge condition group averaged scores of 6.0 and the control group members averaged 5.75, which was significantly different, $p < .001$.

H4 (badge group participants will change strategies less than control group participants) was supported. A *t*-test showed a significant difference between the two groups toggling among different strategies. Participants with access to badges changed strategies significantly less often (just under two times per scenario, on average) than those without access to badges (4 times per scenario), $p < 0.00003$.

The results for H3 and H4 were as expected, although badge group participants were more confident in answers that were less correct, meaning they chose the top-ranked option less often than did participants in the control group. The results for H4 are particularly striking: the control group members changed ranking strategies twice as much as did the badge group members. The badge group members did not have to toggle among the weighting strategies to see which options were top-ranked.

The results for H1 and H2 were not as expected, so we looked closer at how we had conducted the experiment. We found a difference between the two participant groups that we think was likely to have confounded the results. Almost all of the participants in the control group learned English as a first language, whereas most of the participants in the badge group learned English later in life. We noticed that control group participants needed much less time to finish reading the training materials than the badge group, indicating that there may be differences in reading fluency and comprehension between the two groups that may have affected performance.

Nevertheless we learned some things from the empirical testing. For example, we found that people may have different mental models of what a “weighting strategy” means. Based on feedback that we received from the badge group, some people felt the “emphasize maximum” badge should be associated with the box-plot with the highest maximum. This is not the case. The “emphasize maximum” badge indicates the option whose box-plot has the lowest (best) maximum value (although the other box-plot components also play into this comparison to a lesser extent than the maximum value). This misunderstanding tells us that we need to rename the strategies.

We also believe that our placement of the badge legend may have introduced some confusion. Some participants were not sure whether the badge legend linked the meaning of the badges to weighting strategies or to box-plot elements (that is, maximum, $75^{th}$ percentile, median, $25^{th}$ percentile, and minimum).

## FUTURE WORK

The results from H3 and H4 are encouraging enough that we feel it is reasonable to continue refining the badge concept. An obvious next step would be to revise the information presentation to have a clearer legend and better names for the weighting strategies, then rerun the badge condition with participants whose demographic characteristics more closely matched those of the control condition.

If the new experimentation shows that the badge condition improves performance, it would be worthwhile to embark upon additional testing using an experiment participant population consisting solely of people with emergency response experience. Research performed to date has concentrated on basic cognitive processes common to all people and so we did not feel the need to screen specifically for emergency response expertise (although some of our participants had this experience). Now that we are investigating user interaction preferences at a more detailed level, however, it is time to recruit emergency responders for experimentation.

As part of future experimentation we plan to develop modifications to the visualization to enable decision makers to dig deeper into the data used to create the box-plots. The idea is to show the factors that cause cases to have better or worse outcomes. An additional simple modification would be to show an up or down arrow next to a rank to depict whether an option's rank has changed since the last weighting strategy was selected.

## ACKNOWLEDGMENTS

We would like to acknowledge the assistance provided by Mark S. Pfaff and Yikun Liu of Indiana University Indianapolis and Gary L. Klein of The MITRE Corporation. We also thank the experiment participants.

## REFERENCES

1. Bankes, S. C. (1993), Exploratory Modeling for Policy Analysis, *Operations Research*, 41, 3, 435–449.
2. Campbell, J. L. (2004), *In-Vehicle Display Icons and Other Information Elements: Volume II: Final Report*, Office of Safety and Traffic Operations Research and Development, Federal Highway Administration, McLean, VA.
3. Chandrasekaran, B. and Goldman, M. (2007), Exploring Robustness of Plans for Simulation-Based Course of Action Planning: A Framework and an Example, *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making*, Honolulu, HI.
4. Christ, R. E. (1984), Research for Evaluating Visual Display Codes: an Emphasis on Colour Coding, in R. Easterby & H. Zwaga (Ed.), *Information Design: The Design and Evaluation of Signs and Printed Material*, John Wiley, New York, NY, 209 – 228.
5. Cleveland, W. S. and McGill, R. (1995), Graphical Perception and Graphical Methods for Analyzing Scientific Data, *Science*, 229, 828 – 833.
6. Conati C. and Maclaren, H. (2008), Exploring the role of individual differences in information visualization, in *Proc. Advanced Visual Interfaces (AVI)*, 199–206.
7. Drury, J. L., Klein, G. L., Pfaff, M., and More, L. (2009), Dynamic Decision Support for Emergency Responders, *Proceedings of the 2009 IEEE Technologies for Homeland Security Conference*, Waltham, MA.
8. Gupta, P. and Pu, P. (2003), Social Cues and Awareness for Recommendation Systems, *Proceedings of the 2003 International Conference on Intelligent User Interfaces*, Miami, FL.
9. Keeney, R., and Raiffa, H. (1993), *Decisions with Multiple Objectives: Preference and Value Tradeoffs*, Wiley, New York, NY.
10. Klein, G. A. (1999), *Sources of Power: How People Make Decisions*, MIT Press, Cambridge, MA.
11. Klein, G. L., Pfaff, M. and Drury, J. L. (2009), Supporting a robust decision space, *Proc. of the 2009 AAAI Spring Symposium on Technosocial Predictive Analytics*, Stanford U., Palo Alto, CA.
12. Mackinlay, J. (1986), Automating the design of graphical presentations of relational information, *Trans. On Graphics*, 5(2), 110 – 141.
13. Norman, D. (1988), *The Design of Everyday Things*, Basic Books, New York, NY.
14. Nowell L., Schulman, R., Hix, D. (2002), Graphical Encoding for Information Visualization: An Empirical Study, *Proc. of the IEEE Symposium on Information Visualization (InfoVis'02).*
15. Pfaff, M., Drury, J. L., Klein, G. L., and More, L. (2010a), Decision Support for Option Awareness in Complex Emergency Scenarios, *Proceedings of the Applied Human Factors and Ergonomics Conference*, Miami, FL.
16. Pfaff, M., Drury, J. L, Klein, G. L., More, L., Moon, S. P., and Liu, Y. (2010b), Weighing Decisions: Aiding Emergency Response Decision Making via Option Awareness, *Proceedings of the 2010 IEEE International Conference on Technologies for Homeland Security*, Waltham, MA.
17. Tukey, J. W. (1977), *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.