Evaluating the Effectiveness of the PreMortem Technique on Plan Confidence

Beth Veinott
Klein Associates Division
Applied Research Associates
bveinott@ara.com

Gary A. Klein
Klein Associates Division
Applied Research Associates
gklein@ara.com

Sterling Wiggins
Klein Associates Division
Applied Research Associates
swiggins@ara.com

ABSTRACT

One problem affecting crisis management planning teams is overconfidence– an inflated belief that a plan will be successful. In this paper we compared the effect of several different methods for reducing individual team member confidence levels and compared each to a baseline control condition. One hundred and seventy-eight people participated in one of five conditions to evaluate an H1N1 flu epidemic plan in a university context. Over the course of evaluating the plan, participants provided several ratings of confidence in the plan's success and their understanding. We compared several techniques commonly used, such as critique, Pro/Cons generation, Cons only generation and a newer technique, PreMortem, to a baseline condition. The Pro/Cons generation, Cons only generation and the PreMortem technique all reliably reduced confidence levels more than baseline condition. Furthermore, the Premortem method, imagining that a plan has failed and then generating reasons to explain why, reliably reduced confidence more than each of the other conditions, and therefore can be a useful tool for combating overconfidence in crisis management planning. We discuss the results in the context of sensemaking and decision making theory.

Keywords:

Planning tools, Plan Evaluation Techniques, Sensemaking

INTRODUCTION

Effective crisis management response depends on good planning (Gheytanchi et al., 2007). Evaluating plans is essential for their success, but also very difficult (Chermack, 2006). It is essential because complex situations harbor all kinds of complications and risks, several of which were highlighted by Gheytanchi and colleagues regarding the Katrina response. Some of them can be identified or at least suspected in advance, whereas others will not be discovered until after the plan is being executed. Therefore, the team or organization needs to anticipate potential flaws in a plan prior to execution, and to create a culture in which unanticipated flaws are quickly spotted and managed.

The critiquing process is difficult for many reasons. Some of these reasons are that the team or organization may lack the experience to spot flaws. It may not have enough time to conduct a serious review of the plan. It may be so confident that it doesn't see a need for a critique. The members of the team/organization may be reluctant to speak out for fear of disrupting harmony and marginalizing themselves.

These barriers are serious for any type of team/organization, but they are particularly threatening to crisis management operations (Andersson, Pilemam, and Hallberg., 2008). Leifler (2008) has proposed a ComPlan (Combined Collaborative Command and Control Planning) tool to help planners get feedback on constraint

Reviewing Statement: This paper has been fully double blind peer reviewed.

violations and synchronization needs, illustrating the importance of critiquing support. Life-Saver (Sabino et al., 2008) is an agent-based simulation system for dam-break emergencies. WIPER (Wireless Phone-based Emergency Response) (Schoenharl et al., 2006) is a multi-agent system for emergency response that is intended to help planners evaluate possible courses of action. While tools such as ComPlan, Life-Saver and WIPER can augment a team's capabilities in certain areas, teams still need methods for evaluating plans. Crisis management presents challenges of situational complexity and ill defined problems that information technology tools won't handle as flexibly as a human does. Therefore developing a planning tool kit that includes both technological solutions and team strategies seems most promising for improving crisis management planning.

Turoff et al. (2006) describe emergency preparedness gaming as a way to help crisis teams identify essential factors that might not be considered. The use of gaming is extremely valuable for getting teams up to speed. Our interest is identifying actions that teams can take in an emergency situation when they need to be candid about problems and plan weaknesses. These teams have to offer blunt criticisms without damaging relationships. They have to temper overconfidence without harming morale.

There are several methods that people use when evaluating a plan and that have been offered as ways to improve the planning process. Interestingly, there is very little research comparing different plan evaluation techniques. For example, teams are often asked to "critique" a plan. Red Teams are used to provide outside evaluations. However, the logistics and time required for assembling a Red Team, conducting the review, considering the recommendations, and making the necessary changes may be impractical in many emergency settings. Devil's Advocates have been proposed – identifying a team member who is responsible for taking a contrarian position. Unfortunately, research suggests that Devil's Advocates are ineffective in improving plan quality (Nemeth et al., 2001). The Devil's Advocate is not treated seriously because the team members assume the criticisms are inauthentic. Even worse, the team may relax its own skepticism because the critiquing function has been outsourced to the Devil's Advocate. Finally, Benjamin Franklin advocated making decisions by developing pros and cons associated with each course of action in order to determine the best course of action (Cozier, 1978; Janis and Mann, 1977). Each of these methods can facilitate plan evaluation and improvement in theory, but the relative effect on confidence remains to be determined.

In addition to the methods listed above, more recently, the PreMortem method has been proposed as an effective plan evaluation technique (Klein 1998; 2007). For example, a post-mortem refers to an autopsy conducted on a patient to determine the cause of death. The autopsy provides useful information to the medical community and to the family but it does nothing to help the patient. The PreMortem method moves the "autopsy" forward. Thus, a project team may include a PreMortem during the Kick-Off meeting. A PreMortem is an exercise conducted around the premise that a program or plan or project has failed. The team generates plausible reasons for this failure, and these reasons become the plan critique.

Step 1: Prepare by getting familiar with the plan
Step 2: Imagine the plan completely failed – a fiasco
Step 3: Generate reasons for failure
Step 4: Consolidate the lists of reasons
Step 5: Revisit the plan: What can be done to prevent the fiasco?

Table 1. Steps in the PreMortem Process

Table 1 shows the steps of a PreMortem. First, the team members become familiar with the plan. Then they are told that an infallible crystal ball shows that this plan was a fiasco. The time scale in the crystal ball (e.g., six months in the future) clearly reveals that the plan didn't work. But it doesn't show why. That's the job of the team. The team members are given a few minutes to individually write down all the reasons they can think of regarding why the plan has failed. Third, the facilitator has each member announce what was on his/her list, starting with the team leader and going around the table getting one reason from each person, then going around for a second sweep and, if necessary, additional sweeps. Because this step appears to greatly reduce confidence in the plan, the fourth step asks the team members to scan the list and each suggest one action that they will add that could help to reduce the likelihood of the fiasco pictured in the crystal ball. The final steps involve revising the plan in light of the weaknesses, and then periodically revisiting the PreMortem list to see if any of the items are appearing.

Why might the PreMortem reduce overconfidence? About the time that the PreMortem method was being developed, Mitchell and colleagues published a study that shed some light on why the PreMortem method might be effective (Mitchell, Russo, Pennington, 1989). They asked participants in a controlled study to imagine the outcome of a future event (not a plan). Setting the perspective in the future or the past had no effect on the number of explanations, but uncertainty did have a strong effect. Participants told that the event's occurrence was certain generated more reasons than participants who were told it was uncertain. This "prospective hindsight" effect increased the number of reasons for the event's occurrence by approximately 30%. We hypothesize that this prospective hindsight is part of the PreMortem method could reduce confidence.

The PreMortem method takes advantage of this prospective hindsight effect by insisting on the certainty that the fiasco has occurred, so there is no energy wasted in quibbling about likelihoods. The PreMortem method differs from a traditional critique in several other ways. The team members work individually for two minutes, to increase efficiency by getting independent viewpoints. Also, a traditional critique poses barriers to offering criticisms because the members are criticizing a plan that the leader has put forward, and which they themselves may have helped develop. If the plan is flawed, then perhaps the planners were not very good. In contrast, during a PreMortem exercise the way the team members show they are smart is by generating novel and insightful criticisms. There is a healthy competition going on. In addition, the exercise is preparing the team to be self-critical as they implement the plan.

Based on the comments of the participants themselves, there has been a long held belief that the PreMortem reduces confidence and can improve confidence calibration. Overconfidence appears to be an important problem for many teams, and a number of researchers have attempted to understand the underlying cognitive mechanisms (c.f., Koehler, 1991 for a review; Klayman, Soll, Gonzalez -Vallejo, & Barlas, 1999; Koriat, Lichtenstein, & Fischhoff, 1980; Lichtenstein & Fischhoff, 1977). In a study evaluating people's average confidence ratings relative to their average accuracy, Koriat et al. (1980) evaluated the hypothesis that people's confidence in their judgment is determined by the type of reasons that they generate. Consistent with this hypothesis, they found that people who generated only reasons in favor of their judgment were more overconfident (i.e., more confident on average than accurate) than people who generated reasons both why their judgment might be wrong and why it might be correct. A similar explanation (only generating reasons in favor of an estimate) has been put forth as an explanation for the planning fallacy, the notion that people tend to underestimate task completion times (Buehler, Griffin, & Ross, 1994).

We conducted an experiment to evaluate the effectiveness of several different methods (Baseline, Critique, ProCons, Cons Only, PreMortem) on plan confidence. We were particularly interested in whether the PreMortem reduces confidence because it has people focus on reasons why the plan failed. In doing so, we compared this method to a baseline and three other methods that people typically use in groups and (Critique, Pro/Cons, and Cons Only) for a total of five conditions. We did not directly examine overconfidence – because it typically involves estimates based on the proportion of correct answers in a general knowledge type questions. Rather, we obtained estimates of confidence in a plan prior to the experimental manipulations and then subsequent to it. We were interested in the effect of these different methods on confidence, understanding, and plan quality. In this paper, we are reporting only on the confidence and understanding findings.

These methods differ in their evaluation focus (strengths, weaknesses, and failure). In the PreMortem method, there are two mechanisms working together to produce the reduction in confidence. First, people focus on why the plan might fail. Second, people generate concrete reasons and evaluate confidence based on a certain outcome (e.g., assume the plan failed). Other methods such as the Pro/Cons or Cons only involve only one of these mechanisms, not both. Therefore, differences we see between the PreMortem and the Cons or ProCons condition may be attributed to this difference first outlined in Mitchell et al. (1989). In this paper, we evaluate three hypotheses:

> **H1**: The PreMortem method reduces confidence in a plan (relative to initial ratings) more than other methods. We will measure this by evaluating the difference between baseline plan confidence ratings and those collected directly after the experimental conditions.
>
> **H2**: Resolving the problems identified during PreMortem increases confidence more than the other techniques. This hypothesis will be evaluated by comparing the change in confidence over time. This will be accomplished by comparing the ratings immediately after each experimental condition with the confidence in the plan after generating possible solutions for the identified problems.

H3: The PreMortem method increases subjective understanding of the plan and context relative to the other conditions. This will be evaluated by comparing the subjective understanding ratings across conditions.

To evaluate these hypotheses and rule out several alternative hypotheses, we ran five conditions in this experiment. The baseline condition represented no additional processing of the plan. There were four experimental conditions, three (CRITIQUE, PRO-CON, CON only) represented realistic evaluation methods and a more recent fourth strategy, the Premortem method, represents a more recent strategy. The methods differ in the focus of the evaluation (strengths, weaknesses, failures, or a combination) and whether the outcome of the plan is probabilistic (may succeed) or certain as it is in the PreMoretem (assume plan fails).

METHOD

Subjects. One hundred and seventy-eight students (62% female) participated in this study for course credit. The average age was 19 years, and they participated in groups of five to six people. Confidence data was not included in the analysis when it was determined that the instructions were not followed. Consequently, there was one group in the PRO-CON condition whose data were excluded from analysis because the participants did not provide individual initial estimates.

Stimulus Preparation. Each plan was written so that it seemed complete initially, but had some significant flaws that would become clear upon greater scrutiny. We chose an H1N1 flu scenario for this experiment after ten participants provided initial ratings of several plans that university students found very familiar (e.g., traffic congestion/parking on campus, communication during disaster response, and H1N1 flu). A second pilot test with 12 groups piloted the entire experiment procedure with two different plans based on the first pilot test results (communication during disaster response and H1N1 response). Initial average confidence rating for the H1N1 flu plan in our pilot ratings with ten participants was 67%. We felt in order to replicate the situation that we wanted to study, confidence needed to be above 50% in order for there to be room for the confidence to change. In the time that we had for the experiment, there was only time to do one problem. Therefore we chose the H1N1 flu epidemic over communication and information dissemination plan during a natural disaster because of its current relevance for students.

Procedure. Our procedure was as follows and took approximately 35 minutes for each experiment with each experimental manipulation taking about 10 minutes of that time. All participants first read a one page description of a fictional H1N1 lockdown plan that would be instigated only in the case of a rapid spread of the H1N1 flu and provide a way for students and classes to continue using distance learning strategies. The plan had several phases. In the first phase, the lockdown plan was announced by a decision of the university president and senior staff. The announcement would come over a cell phone advisory network, physical media (newspaper) and social media. The announcement of a campus lockdown (meaning that students were supposed to stay home and stay put) came early enough so that faculty, staff, and students would be able to collect relevant papers and materials for working at home. Phase II was to enforce the campus quarantine with security restricting access to central campus only to authorized personnel. Phase III discussed the plan to support the residential students with meals, medicine and course work.

In each condition, participants were instructed to read the initial plan and then provide their initial ratings of confidence by answering the following question. *Based on the information you have seen and using the scale below, how confident (0-100) are you that this plan will be effective in containing the H1N1 Flu virus at this university?* With participants being randomly assigned to a condition, there should be no reliable differences in these initial ratings across all five groups and that is what we found $F(4,173)=1.44$, n.s. The means are plotted in Figure 2 at Conf 1.

Next for each condition, participants completed one of five evaluation procedures. In order to control for the effect of time spent thinking about the problem potentially affecting people's confidence ratings, participants in the BASELINE condition completed a filler task while the experimental groups completed their evaluation procedure. In the CRITIQUE condition, participants spent two minutes critiquing the plan and writing their reactions down. In PREMORTEM condition, participants were instructed to assume the plan had failed, and then spent two minutes writing down explanations for why it had failed. In the PRO-CON group, participants thought about the plan for two minutes and then generated a list of pros and cons (strengths and weaknesses) regarding the plan. In the CONS ONLY, they followed the same procedure as the PRO-CON group, but only listed problems with the plan. Examples

of the reasons participants generated for why the plan failed included not having enough personnel and money to continue the lockdown, people finding ways to get on campus after the lockdown, and parent protests.

Following this type of generation, each of the groups (except the BASELINE group who had not generated a list), consolidated their lists as a group. Next each person in each group made their second confidence ratings as well as a rating of understanding. Then each person spent another two minutes generating improvements based on the problems identified with the plan (except the BASELINE group who was instructed to identify ways to improve the plan). Then each group generated a group list of improvements to the plan that the facilitator wrote on the board. Finally, each person provided a final confidence rating, as well as providing standard demographic information.

RESULTS

In Figure 1, the average confidence ratings by condition at Time 1 (initial rating), Time 2 (directly post manipulation) and Time 3 (after generating ways to improve the plan) are all plotted. As can be seen in Figure 1, there are no reliable differences at time1 ($F(4,173)=1.44$, n.s.). From this picture, it is also clear that the baseline group did not have much movement in their confidence ratings either down or up. The critique group also had very little change in their confidence ratings between the Time 1 and Time 2 either.

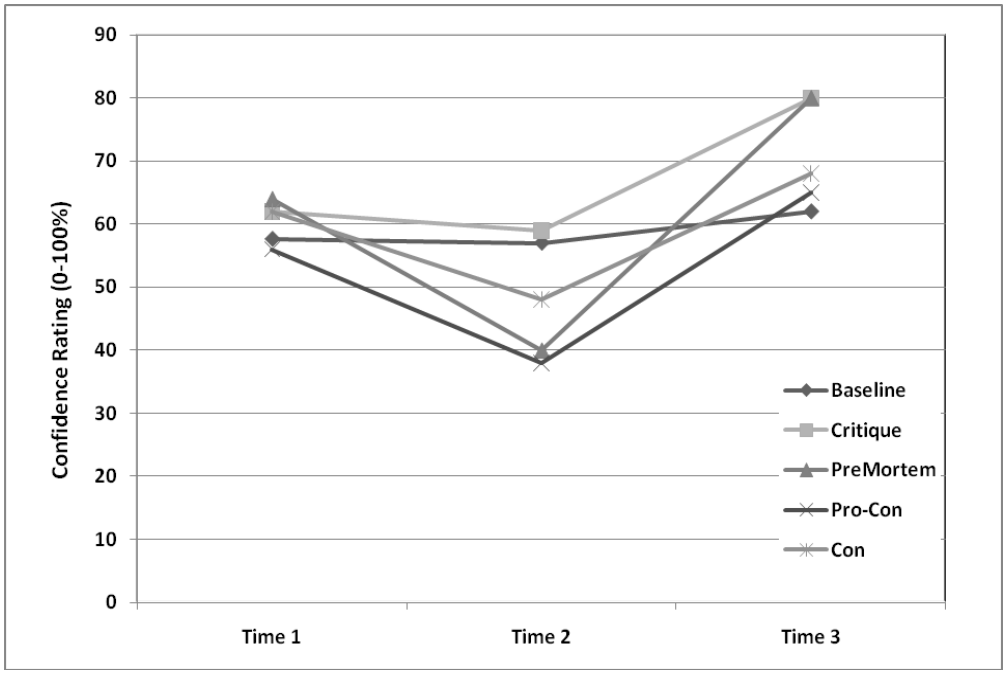

Figure 1. Confidence ratings by condition across three time ratings

While Figure 1 is a plot of the average raw data, Figure 2 plots the change in confidence in order to evaluate the first two hypotheses. The dark colored bars in Figure 2 represent the reduction in confidence between Time 1 and Time 2 for each condition (H1). The light bars represent the increase in confidence between Time 2 and Time 3 for each condition (H2). Next, we will evaluate and report the findings for each hypothesis.

H1: The hypothesis that the PreMortem method would reduce confidence in a plan (relative to initial ratings) more than other methods was supported. By evaluating the difference between confidence ratings at time 1 and time 2, we can determine if the confidence is reduced (dark bars in Figure 2) and whether there is a main effect of condition on confidence ratings. A one-way ANOVA showed a reliable main effect of condition on changes in ratings of confidence ($F(5, 173) = 30.88$, $p < .0001$). As can be seen in Figure 2, the PreMortem ($M= -25.0$, $t(173)= -10.93$, $p <.0001$), PRO-CON ($M= -14.0$, $t(173)= -4.04$, $p <.01$)and CON only methods ($M= -12.4$, $t(173)= -4.19$, $p <.01$) all reduced confidence more than the BASELINE (no reduction) or CRITIQUE conditions. Furthermore, consistent with H1, planned pair wise comparisons revealed that the PreMortem reduced confidence more all the other conditions, including the PRO-CON ($t(74)= -7.08$, $p <.05$) and the CON only group. ($t(67)= -8.42$, $p <.05$).

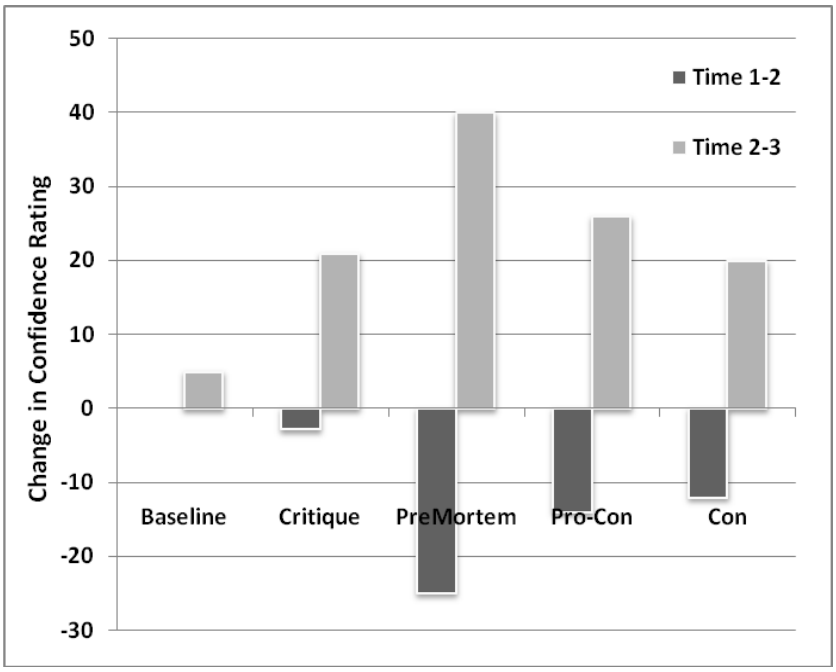

Figure 2. Effect of different methods on confidence (Time 1-2) and resulting solutions (Time 2-3) across conditions (Note: Negative is a reduction in confidence)

H2: The hypothesis that resolving the problems identified during PreMortem increases confidence more than the other techniques was also supported (light bars, Figure 2). This was evaluated by comparing the ratings immediately after each experimental condition (Time 2) with the confidence in the plan after generating possible solutions for the identified problems (Time 3). Consistent with H2, a one way ANOVA produced a main effect of condition on confidence ratings ((F(5, 173) = 72.54.88, $p < .0001$). Planned pairwise comparisons revealed that the PreMortem group had a greater increase in confidence between Time 2 and Time 3 relative to the all the other conditions, including the PRO-CON condition ($t(74) = -4.42$, $p < .05$) which had the second greatest increase. As can be seen in Figure 2, the BASELINE group increased the least.

H3: The hypothesis that the PreMortem method increased subjective understanding of the plan and context relative to the other conditions was partially supported (Figure 3). While a one way ANOVA indicated no main effect of condition on understanding ratings ($F(5,173) = 1.92$, n.s.). Planned pairwise comparisons revealed that the PreMortem ($t(74) = 1.97$. $p < >05$), PRO-CON ($t(67) = 2.02$, $p < .05$) and CON only ($t(54) = 1.88$, $p < .05$) methods all increased understanding relative to the baseline condition, but not the Critique condition. One interpretation of these data is that if one's confidence is not reduced, it may be more difficult to improve the plan and in turn improve one's understanding.

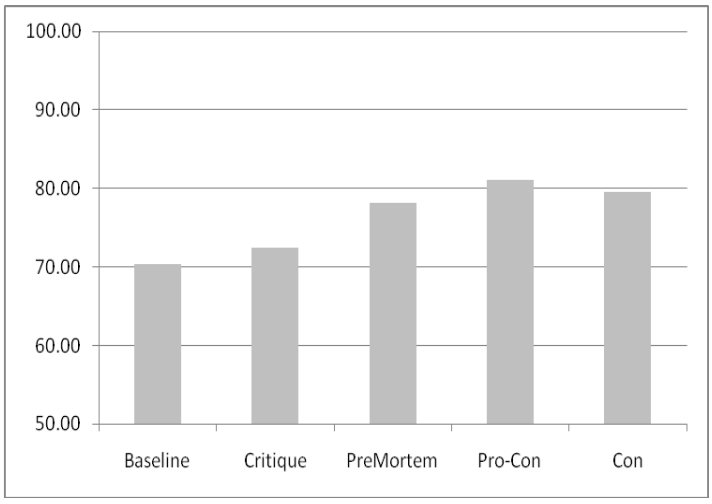

Figure 3. Ratings of Understanding by Conditions

DISCUSSON

In this experiment we set out to evaluate several different lightweight methods for evaluating plans. The results show that the PreMortem method was effective in reducing confidence in the plan described to the participants and in improving understanding to a lesser degree. These findings suggest that the PreMortem method can be a valuable adjunct to crisis management planning and execution. The method is an effective way to evaluate plans prior to implementation. It is also an efficient way to do evaluations during the time pressures of execution.

The PreMortem was found to be better than a host of standard evaluation methods for affecting confidence. The PreMoretem method reduced confidence more than every other technique. In fact, the change in confidence was almost twice that of the PRO-CON group and the CON only group suggesting that both mechanisms, certainty of outcome and a focus on failure, are contributing to confidence estimates. Furthermore, in support of H2, confidence ratings improved most for the PreMortem group between Time 2 and Time 3 relative to the other conditions. This may be due to that group having the lowest average confidence rating at Time 2, so the greatest opportunity for improvement. While we will need to evaluate the reasons that people generated to understand this, one plausible explanation of this effect is that they generated better reasons for failure and consequently better solutions in the PreMortem condition relative to the other conditions.

There are a number of qualifications for our findings. First, the participants were primarily college sophomores, whereas the PreMortem is always used with experienced planners and leaders. We were worried that the effect might not hold with inexperienced participants, but it did. These results suggest it could be even stronger for experienced planners but this will require additional research.

Second, all the experimental groups (CRITIQUE, PRO-CON, CON only) conducted an individual plan review, writing down issues prior to a group discussion. This is not how critiquing sessions are typically conducted. It is, however, the way the PreMortem sessions are run. We needed to incorporate this aspect of a PreMortem into the other conditions in other to maintain experimental control and rule that out as an alternative explanation. To the extent that an individual review time is valuable, we have artificially inflated the quality of the other group sessions (except the BASELINE group).

Third, the participants all reacted to a ready-made plan, whereas teams typically evaluate plans it has, itself, generated. We expect that the method will be more useful when the team is critiquing its own plans, but in order to conduct the controlled experiment we chose to rely on ready-made plans.

Fourth, the PreMortem method is also intended for teams that will have to implement the plan, whereas the participants in our study had no responsibility for implementation. We expect that the PreMortem method will work better when the team members know they are responsible for implementation.

Future research will be needed to address these issues – using experienced planners and leaders, working with plans they have generated themselves and will be preparing to implement. Additional research might compare this method

to other methods (e.g., brainstorming) that are typically used by groups. Another research question addresses the quality of the outputs – how do they affect people's confidence? If our group findings are consistent with Koriat et al.'s (1980) findings, then we would expect that the number of reasons that people generate, the content of those reasons and the difficulty in generating them could affect confidence. Are the first five items the most important ones? Is there value in continuing the generation phase beyond two minutes?

This research also enabled us to think more deeply about various facets of the PreMortem. The focus of a PreMortem session can be on the outcome ("why has the plan failed?") or on weaknesses in the plan itself, or on limitations of the team that is going to carry it out, or on vagaries of the situation that might compromise the plan.

Lastly, we return to the issue of how the PreMortem might be working. One possibility is that it encourages team members to consider alternative scenarios, and thereby confers some of the benefits of planning and decision scenarios (Wack, 1985). The PreMortem may be tapping into some of the factors that may make scenario planning useful. However, the PreMortem technique does not focus on identifying key uncertainties and identifying information about how the future might unfold as is the case in scenario planning. So it is not itself a scenario, nor does the technique instruct participants to think about any specific scenario.

It seems more likely that the PreMortem technique is operating by changing the participants' mental model for evaluating the plan from one where the plan is possibly successful to one where it definitely fails. Choosing a frame is part of the sensemaking process (Weick, 1995) and is directly related to the data/frame model of sensemaking (Klein, Phillips, Rall, & Peluso, 2007). In most situations, people rely on simple pattern-matching to identify an appropriate mental model (i.e., frame) for understanding situations. Sensemaking, a conscious and deliberate process, often begins when people encounter some contrary evidence or reason (i.e., data) to question a frame. Their response may be to hold on to their initial frame, either by explaining the data away or by elaborating the initial frame to fit the new data. A second reaction to contrary evidence is to replace the frame to fit the new data. In many instances, people try to preserve their initial frame, the first of the two reactions presented above. The PreMortem exercise counters that tendency. It directs the team members to take seriously the possibility that their initial frame might be mistaken (i.e., by considering reasons why the plan failed). According to the Data/Frame model, this guidance should result in a figure/ground reversal where anomalies and inconsistencies that had previously been inconvenient are now highly relevant as openings to explore what might go wrong. Data that had been relied on might now be distrusted. Different anchors might be considered for re-framing the situation (Tversky & Kahneman, 1981). Different kinds of data might be highlighted. New questions might reveal that key data are missing – data whose absence was not noted before.

Another explanation for why the Premortem method is effective comes from research on the sources of confidence. Kohler (1991) suggested that plausibility judgments are formulated to account for data, and that the easier it is for a hypothesis to account for data the higher its confidence rating. Therefore, by emphasizing the inconsistencies, the PreMortem method may make it harder to generate plausibility judgments resulting in a reduction in confidence ratings. While Koehler's work focused on judgments regarding explanations of an event, the same mechanism may be operating in judging confidence in a plan.

The PreMortem method does not need to postulate an alternative plan to be effective; it seems to work well simply by conjuring up the image of the plan's failure. As we mentioned in the beginning, the Premortem differs from other methods because it enlists both an inverse frame (i.e., that the proposed plan has failed rather than succeeded) for evaluation and Mitchell et al.'s prospective hindsight (e.g., certainty regarding plan outcome). This idea raises the possibility of selectively using different frames for planning and for plan evaluation, a concept that forms the basis of scenario planning. Instead of asking "what could go wrong," action teams can propose frames that enable their members to be more thoughtful and less overconfident. Although more research regarding plan evaluation techniques is needed, this study indicates that the Premortem is an effective and lightweight planning tool that can aid planners both prior to and during crisis situations in evaluating and improving their plans.

ACKNOWLEDGEMENTS

We would like to thank the U.S. Army's Advanced Decision Architectures Collaborative Technology Alliance for sponsoring this research under the Cooperative Agreement DAAD19-01- 2-0009. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory or the U.S. Government. We also thank the reviewers for their helpful comments.

REFERENCES

1. Andersson, D., Pilemam, S. & Hallberg, N. (2008). *Evaluation of crisis management operations using Reconstruction and Exploration.* Paper presented at the 5th International ISCRAM Conference, Washington, DC.
2. Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the planning fallacy: Why people underestimate their tasks completion times. *Journal of Personality and Social Psychology, 67*, 366-381
3. Chermack, T.J., (2006). Assessing the Quality of Scenario in Scenario Planning, *Futures Research Quarterly,* pp. 23-35.
4. Cosier, R. (1978). The effects of three potential aids for making strategic decisions on prediction accuracy, *Organizational Behavior and Human Performance, 22,* pp. 295-306.
5. Gheytanchi, A., Joseph, L., Gierlach, E. Kimpara, S., Housley, J., Franco, Z. & Beutler, L. (2007). The dirty dozen: Twelve failures of the Hurricane Katrina response and how psychology can help. *American Psychologist*, 62(2), 118-130.
6. Janis, I. & Mann, L. (1977). *Decision Making.* Free Press: New York.
7. Klayman, J., Soll, J. B., Gonzalez -Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior & Human Decision Processes, 79*, 216-247.
8. Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.
9. Klein, G. (2007). Performing a project PreMortem. Retrieved September, 2007, from http://harvardbusinessonline.hbsp.harvard.edu/b02/en/common/item_detail.jhtml;jsessionid=YIP1WQM4XBGBEAKRGWCB5VQBKE0YOISW?id=F0709A&referral=7855
10. Klein, G., Phillips, J. Rall, E. and Peluso, D. (2007). A data-frame theory of sensemaking. In R. Hoffman(Ed) *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, (113-155). Lawrence Erlbaum Associates.
11. Koehler, D. (1991). Explanation, Imagination, and Confidence in Judgment. *Psychological Bulletin*, 110, 499-519.
12. Koriat, A., Lichtenstein, S., Fischhoff, B. Reasons for Confidence. *Journal of Experimental Psychololgy: Human Learning and Memory*, 6, 107-118.
13. Leifler, O. (2008). *Combining technical and human-centered strategies for decision support in command and control: The ComPlan approach*. Paper presented at the 5th International ISCRAM Conference, Washington, DC.
14. Lichtenstein, S. & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgements. *Organizational Behavior and Humand Performance*, 20, 159-183.
15. Mitchell, D. J., Russo, J. E., & Pennington, N. (1989). Back to the future: Temporal perspective in the explanation of events. *Journal of Behavioral Decision Making, 2*, 25-38.
16. Nemeth, C., Rogers, J., & Brown, K. (2001). Devil's advocate vs. authentic dissent: Stimulating quantity and quality. *European Journal of Social Psychology, 31*, 707-720.
17. Sabino, A., Nóbrega, R., Rodrigues, A., & Correia, N. (2008). *Life-saver: Flood emergency simulator.* Paper presented at the 5th International ISCRAM Conference, Washington, DC.
18. Schoenharl, T., Madey, G., Szabó, G., & Barabási, A. L. (2006). *WIPER: A multi-agent system for emergency response.* Paper presented at the 3rd International ISCRAM Conference, Newark, NJ.
19. Turoff, M., Chumer, M., & Hiltz, S. R. (2006). *Emergency planning as a continuous game.* Paper presented at the 3rd International ISCRAM Conference, Newark, NJ.
20. Tversky, A. & Kahneman, D. (1981). The framing and the psychology of choice. *Science,* Vol 453, 252-258.
21. Wack, P. (1985), Scenarios: Uncharted waters ahead. *Harvard Business Review,* 63, 72-89.