Supporting Decision Making by a Critical Thinking Tool

Article *in* Human Factors and Ergonomics Society Annual Meeting Proceedings · September 2005
DOI: 10.1177/154193120504900364

Experimental Evaluation of a Critical Thinking Tool to Support Decision Making in Crisis Situations

Jan Maarten Schraagen, Aletta Eikelboom, Kees van Dongen and Guido te Brake
TNO Human Factors, Soesterberg, The Netherlands
{schraagen, eikelboom, vandongen, tebrake}@tm.tno.nl

ABSTRACT

Building up proper situation awareness is one of the most difficult tasks in the beginning stages of large-scale accidents. As ambiguous information about the events becomes available, decision makers are often tempted to quickly choose a particular story to explain the events. Subsequent information that contradicts the initial story may easily be discarded and cognitive tunnel vision takes over. Our approach, as part of the COMBINED Systems project, is to prevent tunnel vision by providing critical thinking support. In a laboratory experiment with 60 participants, we tested this hypothesis by comparing the Critical Thinking tool with a 'no support' control condition and a 'minimal support' condition. Participants acted as crisis managers determining the likely cause of an accident based on different pieces of information. The results show a positive impact of the tool on both the decision making process and decision making effectiveness.

Keywords

Critical thinking, decision support, crisis management.

INTRODUCTION

Extensive research in disaster crisis management (Quarantelli, 1989; 1999) has shown that the source and locus of most problems in the emergency phase of disasters is *not* to be found in the victims but in the organizations attempting to help them. The organizations involved struggle with problems of communication and information flow, authority and decision making, and coordination. Building up proper situation awareness is one of the most difficult tasks in the beginning stages of large-scale accidents. In novel or ambiguous situations, which are often found in large-scale accidents, the pattern-matching and recognition-primed strategies of experts (Klein, 1998) can often not be successfully applied for building up situation awareness. This is because the relevant cues about the properties of the accident that are needed to activate memories about emergency response plans are missing, or because knowledge about the proper emergency response plans themselves is missing.

In ambiguous situations, missing information is usually inferred by constructing a story around existing information (Pennington & Hastie, 1993). Although this generally is an effective strategy, there is a danger that people tend to stick with their initial explanation of an ambiguous situation, even when information contradicting their initial interpretation is accumulating as the situation unfolds (framing bias). They also tend to use neutral information as confirmative to strengthen their initial explanation. A final decision bias is that people tend to focus on information that confirms their initial explanation (confirmation bias) and easily discard and forget information that contradicts their previously formed explanation.

Going beyond descriptive theories of expert decision making, Cohen, Freeman and Wolf (1996) have proposed a prescriptive theory, which intends to improve expert decision making in novel and ambiguous situations. Based on critical incident interviews with active-duty naval officers, they concluded that proficient decision makers are not only recognitionally skilled, as for instance described by Klein's theory (1998) of expert decision making, but also meta-recognitionally skilled. Proficient decision-makers are able to recognize when situations are unfamiliar and problematic and use this knowledge to both initiate story-construction and critical thinking strategies so as to improve situation awareness.

Critical thinking

We think that situation assessment in complex, dynamic and ambiguous domains, such as the crisis domain, can be supported with critical thinking support tools. We think that problems with situation assessment, such as framing and confirmation biases, can be reduced with critical thinking support. Before we describe how critical thinking can be improved by visualization and critiquing we need to understand what critical thinking is.

We think critically when we ask ourselves questions, such as: How do I know this? Are the reasons for this belief adequate or do I miss crucial information? What situations can I think of in which this belief is false? Why did I decide to do this? Will this action achieve the indented effect? What situations can I think of in which this plan is failing? Is there something else I need to think about? What are the arguments against this belief or course of action? Is there a better belief or plan? What is my real purpose, and am I addressing the issues that really concern me? In the literature on informal logic, probability theory, decision theory, cognitive psychology, communication theory, and rhetoric, there are many ways in which critical thinking has been defined. An important common element in these definitions is that critical thinking involves the deliberate evaluation of intellectual products in terms of appropriate standard of adequacy (Cohen, Salas & Riedel, 2002). Cohen, Freeman and Wolf (1996) have proposed a cognitive theory that prescribes how this can be done.

Cohen et al.'s model of critical thinking describes the interaction between cognitive and meta-cognitive processes. In their model a meta-cognitive process, called quick-test, is used to determine whether it is worthwhile to think more about a situation. If this is the case cognitive processes are used to identify evidence-conclusion relations within the story; critique the story for incompleteness, conflict and unreliability; and attempt to improve it by collecting or retrieving new information and revising assumptions.

In more detail, the quick-test is used to assess whether the situation is unfamiliar, the cost of delay of extra thinking is acceptable and the cost of an error is high. If this is the case, decisions about a course of action are inhibited and evaluative processes are initiated. According to their model three kinds of problems with story telling need to be assessed so as to assess its adequacy: incompleteness, conflict, and unreliability. In case of incompleteness, evidence for a hypothesis is missing. If there is a conflict, arguments provide support both for and against a conclusion. In case of unreliability, the relation between evidence and conclusions is based on unsupported assumptions. If these problems are detected, the story is judged to be inadequate and cognitive regulation processes are used to repair the story. In this process additional data can be collected, attention can be shifted to relevant knowledge, or assumptions can be added or dropped.

Support tool

The critical thinking tool we envision should make the operator aware of incompleteness, conflict and unreliability. We think that by visualizing the argument structure, that is the relation between evidence and hypotheses in the interface of this tool, we can reduce problems with situation assessment such as story building (framing) and confirmation biases. In the first phase of the COMBINED SYSTEMS project, we have investigated how conflict and unreliability can be visualized in the interface of a critical thinking tool and how neutral (non-informative) information can be detected. The critical thinking interface is shown in figure 1.

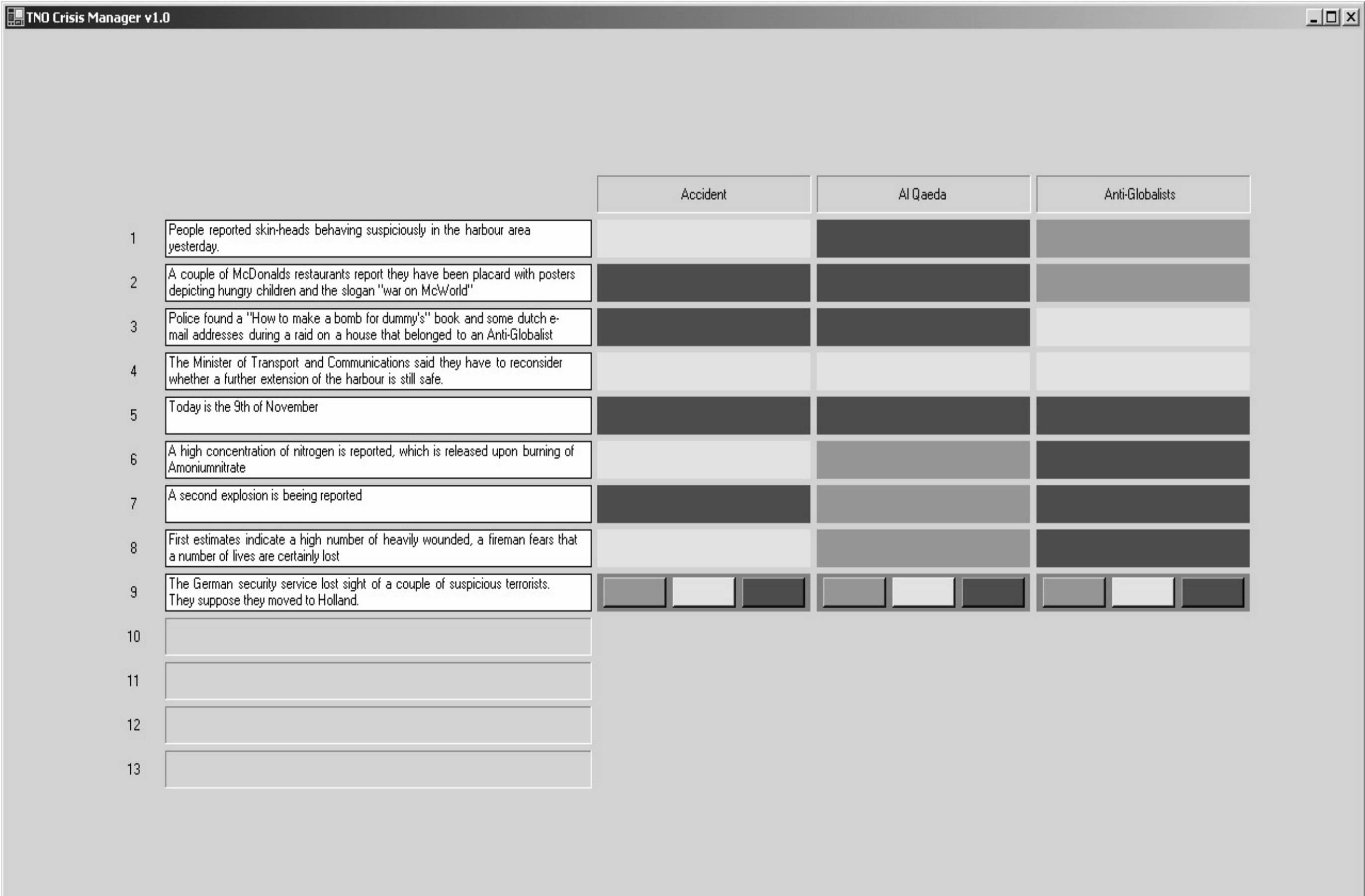

Figure 1. Critical Thinking support tool

We predicted a positive effect of the support tool in terms of number of correct hypotheses chosen, as well as in terms of underlying cognitive processes. We compared the critical thinking tool with two alternatives: (1) a ‘no support’ condition, in which messages have to be interpreted as they arrive, (2) a ‘minimal support’ condition that only lists the likely hypotheses but that does not require any further processing by the participant in the experiment. The critical thinking tool, on the other hand, listed the likely hypotheses and also required the participants to determine for all the messages the level of support for each hypothesis. The level of support is represented by colors: ‘green’ if the particular hypothesis was supported by the message, ‘yellow’ if the hypothesis was partially supported, and ‘red’ if the hypothesis was not supported. The minimal support condition was included to assess the effect that mentioning the hypotheses would have on interpreting the evidence. If full support and minimal support do not differ from each other, we have to conclude that merely mentioning the hypotheses is sufficient to induce critical thinking. If, on the other hand, full support differs from minimal support, we have to conclude that the effect is specifically due to evaluating the evidence in the light of all the hypotheses.

We also investigated the effect of the order in which different messages arrive. Messages could arrive in blocks supporting a particular hypothesis (‘framing’) followed by non-informative messages, or they could arrive in random order (‘no framing’). In the framing condition, participants first received three pieces of information that supported the first hypothesis followed by two non-informative messages. We expected this manipulation to frame the participants in thinking in one direction. After these first five messages, they received four messages supportive for the second and the most likely hypothesis followed by three messages diagnostic for the third hypothesis and one non-informative message. To avoid primacy and recency effects, messages supporting the most likely hypothesis were presented in the middle. We hypothesized that the critical thinking tool would be particularly helpful in the framing condition. In the framing condition, unaided participants are more likely to interpret the two non-informative messages as confirmatory for the first hypothesis due to the story building bias. They are therefore more likely to select the wrong hypothesis. We expect that participants using the critical thinking tool will overcome this bias and are more likely to select the second, correct, hypothesis. The critical thinking tool will result in a more coherent story on the basis of which participants will be able to make more correct judgments regarding the individual messages that support or contradict a particular hypothesis or are non-informative.

METHOD

Participants

Sixty mostly college students (27 male, 33 female) participated in the experiment. Their ages ranged from 17 to 30 years (*M*= 22 yrs). They were paid 40 euros for their participation. Participants were randomly assigned to one of the three experimental groups, but the male/female ratio was kept equal across groups. Being college students, participants were not experienced in crisis management.

Task

Participants acted as crisis managers in the experiment. Their task was to interpret incoming messages containing information following an explosion in a tunnel. The cover story was that the Dutch Prime Minister hosted a meeting with the British Prime Minister and the US president, and that they were en route from Zeeland (a province in the south-west of The Netherlands, and place of birth of the Dutch Prime Minister) to The Hague (seat of Dutch parliament). Suddenly, in one of the tunnels the convoy had to pass through, an explosion was heard. The crisis manager's task was to determine the likely cause of the explosion. For experimental purposes the task was simplified and kept under control by allowing only three possible causes for the explosion: an accident, an attack by Al Qaeda, or an attack by anti-globalists. The crisis managers acted in six independent scenarios. Each scenario contained 13 messages that the crisis manager could view in sequence (the task was self-paced). After reading the 13 messages the crisis manager had to indicate which of the three causes was the most likely in that particular scenario.

Materials

The task was carried out on a computer. The software was developed using Microsoft Visual Basic. All materials, including the training materials, could be read on the computer screen. Background information on Al Qaeda, the anti-globalists and tunnel accidents was also provided on paper during the training phase. The hard copies were removed after participants had read them. A map of the environment where the crisis took place was provided for continuous access.

Procedure

The experimenter welcomed participants. They signed the informed consent form in which they were briefed about the nature of the experiment. Participants then entered the learning phase of the experiment. This phase consisted of the reading of background information on Al Qaeda, the anti-globalists and tunnel accidents. Participants were given information on how Al Qaeda prepares an attack, what kind of chemicals they typically use, and how to recognize an explosion caused by an Al Qaeda bomb. Similar information was provided for the anti-globalists, with particular emphasis on the reasons for committing an attack and the rationale for choosing a location. For tunnel accidents, information was provided on the risks posed by trucks with flammable cargo in tunnels and the consequences when such trucks are involved in accidents (the particular tunnel in the scenarios was not closed to trucks with dangerous goods). Next, participants had to use this information to perform three learning tasks (or games). First they had to link 22 descriptions to given concepts; secondly, they had to increase the likelihood of each of the three hypotheses by selecting particular pieces of information; and thirdly, they had to indicate for 12 pieces of information for all three hypotheses how likely each hypothesis would be given that piece of information; feedback was provided after each piece. For each of these exercises, scores were calculated. At the end of the learning phase, participants were quizzed on their knowledge by a pre-test. The pre-test consisted of a task similar to the experimental task: participants were required to determine for a particular piece of information how likely each hypothesis could explain that information (for instance: the information that the casualty rate is extremely high most likely points to Al Qaeda because they intend to harm as many people as possible; anti-globalists, on the other hand, would try to avoid casualties so this hypothesis is less likely). Participants were required to score at least 85% correct on the pre-test. This percentage was determined on the basis of a pilot study which had shown that a criterion score of 90% was too high, and did not show justice to the learning that participants had accrued. If they failed to reach this criterion, they were obliged to study that part of the learning phase again on which they had received the lowest score. Participants that had not reached the criterion of 85% correct on the pre-test after four attempts were also allowed to enter the experimental phase. The reason for allowing them to enter the experimental phase was because the number of hypotheses correctly chosen did not significantly differ between those who had reached the criterion and those who had not (63% vs. 54% correct, respectively), Mann-Whitney U, p = 0.27. In total, seven participants had not reached the criterion of 85% correct on the pre-test after four attempts (the average score on the pre-test of these seven participants was 83%; the lowest score was 77%).

The experimental phase consisted of six independent scenarios. For each scenario, participants read 13 messages from a variety of sources, some more credible than others (e.g., a piece of information from the police or the Intelligence Service had to be considered as more credible than information from the general public). After reading these messages, participants indicated which of the three causes or hypotheses was the most likely and they had to indicate, on a scale from 0 to 100, their subjective certainty about their choice. They subsequently had to write a press release in which they had to summarize, in their own words, what they thought had happened in the scenario. After writing the press release, they were given the opportunity to change their mind regarding the hypothesis chosen previously. Finally, participants were asked to select from all the messages just presented in the scenario plus five additional messages, not present in the scenario, the five messages they thought supported their choice of hypothesis and five messages they thought did not support their hypothesis. After this, they went on to the next scenario. After all six scenarios were finished participants took a post-test. This test was identical to the pre-test and was included to measure learning or forgetting effects.

The experiment took approximately three hours, depending on the amount of time participants needed for the learning phase, and the amount of time they took for the experimental phase.

Variables

Two independent variables were manipulated: level of support and framing. 'Level of support' consisted of three levels: (1) no support: participants only read the messages in sequence; (2) minimal support: a matrix was shown with the messages appearing in the rows and the three hypotheses in the columns; (3) full support: in addition to the minimal support matrix, the individual cells could now be filled in by the participants by choosing one of three colors. "Level of support' was a between-subjects variable. 'Framing' was a within-subjects variable and consisted of two levels: (1) no framing: messages appeared in random, though fixed, order; (2) framing: messages supportive of a particular hypothesis (e.g. Al Qaeda) appeared in one block after each other.

The following dependent variables were measured: scores on pre-test and post-test; average time to read a single message; average time to reach a decision as to which hypothesis is correct; average number of correct hypotheses chosen; average subjective level of confidence in the decision taken; average number of correct, incorrect and neutral messages chosen as supportive for the hypothesis.

Experimental design

A 2 (framing; within) * 3 (support; between) * 2 (order; between) design with scenario as nested (repeating) factor was chosen. In order to control for order effects in the framing condition, framing (F) and no framing (NF) scenarios were presented in two different orders: (1) F-NF-F-NF-F-NF; (2) NF-F-NF-F-NF-F. Each participant received a unique identification number. Participants with an odd number received order 1, participants with an even number received order 2.

RESULTS

A first inspection of the data showed large learning effects in the first two scenarios as compared with the last four. On the first two scenarios the average percentage correct was 43%; on scenarios three and four the average percentage correct was 58%; and on the last two scenarios, participants scored on average 68% correct. This increase was significant, $F(2, 108) = 9.65$, $p < .01$. Tukey HSD tests showed that the first two scenarios significantly differed from the last four (1-2 vs. 3-4, $p < .05$; 1-2 vs. 5-6, $p < .01$), which did not differ from each other, $p = .19$. Therefore, the first two scenarios may best be considered as training scenarios in which participants were provided the opportunity to apply their knowledge from the learning phase to the experimental task at hand. For this reason, we decided to base all our remaining results on the last four scenarios only.

Table I provides an overview of the main hypotheses along with the level of support and significance.

Hypothesis	Dependent variable	Result						*p*-value
		Framing			No framing			
		NS	MS	FS	NS	MS	FS	
Better performance for the full support group in the framing condition	Percentage correct hypothesis chosen	60*	60*	80*	58	58	65	* = <.05
Full support group more often chooses	Percentage 'hits': correct messages that	70*	70*	79*	66[a]	69	75[a]	* =.05 a: pairwise

supporting messages	did occur in the scenarios							comparison $p = .09$
Full support group less often chooses non-supporting messages	Percentage 'misses': incorrectly chosen messages	23	22	19	21	20	17	> .20
Full support group chooses fewer non-informative messages to support their choice	Percentage neutral messages	23*	17*	13*	28	20	23	* = <.05
Full support group chooses fewer non-informative messages to contradict their choice	Percentage neutral messages	15[a]	13	5[a]	14[a]	7	12[a]	a: pairwise comparison $p < .05$

Table 1. Main hypotheses and result with level of significance (NS: No Support; MS: Minimal Support; FS: Full Support)

Selective learning or forgetting

On average, participants scored 88.5% correct on the pre-test and 87.5% correct on the post-test. Although this difference is statistically significant, $F(1, 54) = 6.54$, $p < .05$, the score on the post-test does not show a substantial forgetting effect, and remains above the 85% criterion level. More importantly, the decrease from pre-test to post-test does not differ among the three levels of support, $F(2, 54) = 1.11$, $p > .05$. We may therefore conclude that any differences among the three groups cannot be explained by selective learning or forgetting effects.

Percentage correct assessments

Participants in the no support and minimal support conditions scored on average 59% correct hypotheses, while participants in the full support condition scored on average 73% correct hypotheses. This difference was, however, not significant, $F(2,57) = 2.09$, $p > .05$. We predicted better performance for the full support group in the framing condition. The full support group did indeed score 80% correct in the framing condition, while the no support and minimal support groups both scored 60% correct. This difference between the full support group and the no support group was statistically significant on a Kruskal-Wallis ANOVA, $H\ (1, N=40) = 3.93$, $p = .05$. The difference between the minimal support and full support groups was, however, not significant, $H\ (1, N=40) = 2.26$, $p = .13$. The three groups did not statistically differ from each other in the no framing condition, $H\ (2, N=60) = 0.64$, $p > .05$. Figure 2 shows the mean percentage correct in the three conditions for both framing and no framing.

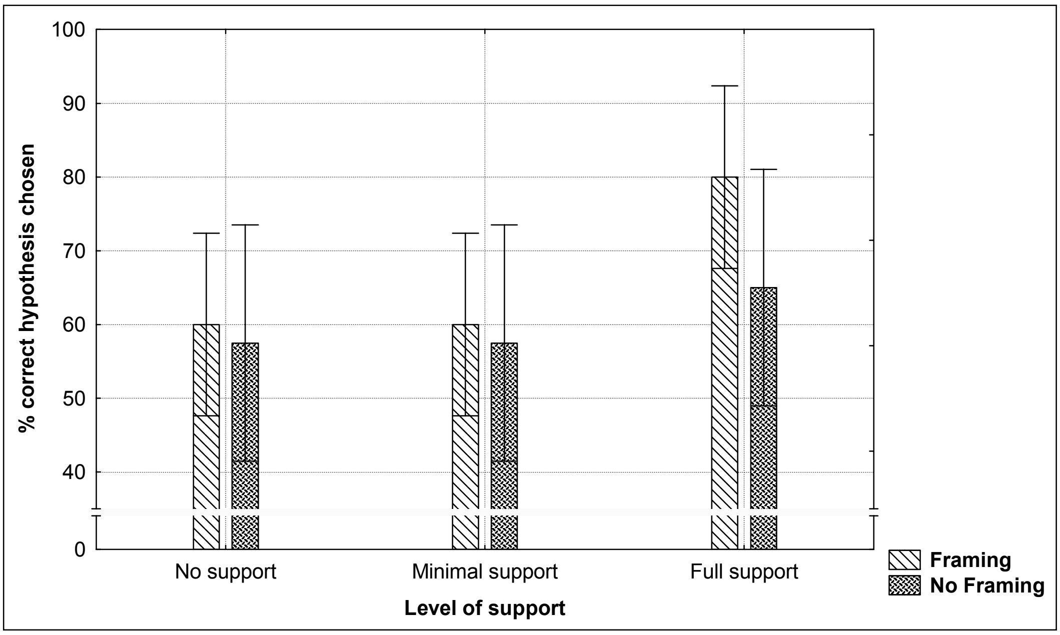

Figure 2. Percentage correct hypothesis chosen as a function of level of support

Assessment errors: confirmation bias

We predicted that participants with full support would more often choose supporting messages and less often non-supportive messages, particularly in the framing condition where participants without the critical thinking tool would be biased. We will first analyze the correct recognition of messages that did indeed occur in the scenarios and supported the hypothesis chosen. A score of 100% indicated that participants had chosen the four correct messages. For the framing condition, participants without any support chose on average 70% correct messages, those with minimal support also chose on average 70% correct messages and those with full support chose on average 79% correct messages. The difference between no support and full support was statistically significant on a Kruskal-Wallis ANOVA, H (1, N=40) = 3.67, p = .05. The difference between minimal support and full support was also statistically significant on a Kruskal-Wallis ANOVA, H (1, N=40) = 4.19, $p < .05$. For the no framing condition, participants without any support chose on average 66% correct messages, those with minimal support chose on average 69% correct messages and those with full support chose on average 75% correct messages. Even though the no framing condition showed the same trend, none of the pairwise comparisons was statistically significant, although the difference between no support and full support approached significance, Kruskal-Wallis ANOVA, H (1, N=40) = 2.82, p = .09.

Next, we looked at the messages incorrectly chosen as being supportive for a particular hypothesis, while in reality they were supportive for another hypothesis. A score of 100% indicated that participants had chosen the five incorrect messages. For the framing condition, participants without any support chose on average 23% wrong messages, those with minimal support chose on average 22% wrong messages and those with full support chose on average 19% wrong messages. There were no significant differences among the three groups (Kruskal-Wallis test H (2, N=60) = 2.04, p > .05). In the no framing condition means were 21%, 20% and 17%, respectively, for the no support, minimal support and full support groups. These means also did not differ significantly from each other, Kruskal-Wallis test H (2, N= 60) = 1.51, p > .05.

Assessment errors: neutral as supportive

We predicted that participants in the full support condition would choose fewer neutral (non-informative) messages to support their choice, especially in the framing condition. The results precisely supported this prediction. For the framing condition, participants without any support chose on average 23% neutral messages, those with minimal support chose on average 17% neutral messages and those with full support chose on average 13% neutral messages. A score of 100% indicated that participants had chosen the three neutral messages. The difference between no support and full support was statistically significant on a Kruskal-Wallis ANOVA, H (1, N=40) = 3.96, p = .05. All other pairwise comparisons were not significant. In the no framing condition, participants without any support chose on average 28% neutral messages, those with minimal support chose on average 20% neutral messages and those with full support chose on average 23% neutral messages. None of the pairwise comparisons in the no framing condition was significant.

Confidence in assessment

As far as level of confidence in the hypothesis chosen is concerned, we predicted that participants in the full support condition would be less confident about the hypothesis chosen, because they would be more aware of alternative hypotheses. The mean level of confidence for the no support group was 60%, for the minimal support group 56%, and for the full support group 61%. This difference was not significant, $F(2,54) = 0.86$, $p > .05$. The critical thinking support tool therefore, does not lead to lower levels of confidence, nor to significantly higher levels.

Assessment time

Average reading time per message was 5.2 s. in the no support condition, 5.9 s. in the minimal support condition, and 7.7 s. in the full support condition (for the full support condition the reading time lasted until participants filled in the first color in the matrix, for the other conditions this was the time until they pushed the 'get-new-message' button). This difference among the three conditions was statistically significant, $F(2,54) = 8.18$, $p < .01$. Tukey HSD tests showed no significant difference between the no support and the minimal support condition, $p > .05$, but they did show significant differences between the full support and no support condition, $p < .01$, and between the full support and minimal support condition, $p < .05$. Total decision time, that is the time from reading the first message until the choice of a hypothesis, was on average 102 s. in the no support condition, 124 s. in the minimal support condition, and 191 s. in the full support condition (this time includes the time needed to fill in the colors for the full support condition). This difference among the three conditions was statistically significant, $F(2,54) = 15.42$, $p < .01$. Tukey HSD tests showed no significant difference between the no support and the minimal support condition, $p > .05$, but they did show significant differences between the full support and no support condition, $p < .01$, and between the full support and minimal support condition, $p < .01$.

DISCUSSION

When confronted with novel or ambiguous situations, expert decision makers usually perform a quick test to decide whether they have enough time to engage in critical thinking and whether the extra time involved with critical thinking outweighs the consequences of an erroneous decision (Cohen et al. 1996). If the quick test tells them not to engage in critical thinking, but rather to decide now, they are prone to make a quick recognition-based decision based on an incomplete, conflicting or unreliable story. They are framed by the available evidence that biases them in a certain direction. Our research has tried to simulate the decision situation experts frequently face. One group of participants did not engage in critical thinking, hence was prone to decision biases when appropriately framed. The other group of participants was stimulated to engage in critical thinking by providing them with a decision support tool that allowed them to repair the incomplete, conflicting or unreliable story. The third group was continuously reminded of the existence of alternative hypotheses, but without asking them to engage in detailed assessment of each piece of evidence.

We hypothesized that such a tool would be most appropriate in situations where decision making could easily go astray, that is, in situations where the initial evidence would point in the wrong direction and subsequent evidence would not be strong enough to get the decision makers on another track. In our experiment, we referred to this situation as the 'framing' condition.

As predicted, we found a large and significant performance improvement (from 60% correct to 80% correct) in the framing condition due to the support tool. The reasons behind this improvement in number of hypotheses correctly chosen may well lie in a more complete and reliable story that was being developed by those using the support tool. This story may be used afterwards when participants were asked to indicate which messages did or did not support their hypothesis. Using a complete and reliable story may have enabled participants to more frequently indicate the correct messages and to less frequently choose non-informative messages. We plan to further corroborate this result by analyzing the 'press releases' that participants wrote after having made their choice of hypothesis. We expect that these stories will also be more complete and reliable in the full support group than in the other groups.

For practical purposes, it is important to note that these advantages of the support tool came with a disadvantage in terms of the time needed to read and interpret the messages. Critically assessing the evidence and building up a coherent story takes time. When time is of the essence, when there is no ambiguity, or when the cost of an error is low, then the support tool should not be used. To save time, we are currently investigating the possibilities of allocating parts of the assessment task to the computer. This will become urgent in real-life cases where hundreds of messages may come in during a disaster and information overload is a constant threat. Our future research will focus on human-system collaboration issues, where different levels of automation may be chosen depending on the information load.

CONCLUSION

This research is a first attempt at developing a decision support tool geared directly at the thought processes of decision makers during crisis management. Starting from well-known decision biases and a recent theory of decision making in ambiguous situations, we developed a relatively simple tool that improved decision making in those situations where it is really needed.

ACKNOWLEDGMENTS

We would like to thank Paul Burghardt and other members of the COMBINED project team for thinking critically with us. Special thanks also to the rest of the TNO team consisting of José Kerstholt, Noortje Jansen, Josine van de Ven, Tjerk de Greef (software development) and Anja Langefeld (statistical advice).

REFERENCES

1. Cohen, M.S., Freeman, J.T. & Wolf, S. (1996). Metarecognition in time-stressed decicion making: Recognizing, critiquing, and correcting. *Human Factors, 38(2)*, 206-219.
2. Cohen, M.S., Salas, E. & Riedel S.L. (2002). Critical thinking: Challenges, possibilities and purpose. *Cognitive Technologies*. Technical Report 02-1.
3. Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.
4. Pennington, N. & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the juror: the psychology of juror decision making* (pp. 192-221). Cambridge: Cambridge University Press.
5. Quarantelli, E.L. (1989). How individuals and groups react during disasters: Planning and managing implications for EMS delivery. *Preliminary Paper # 138*. University of Delaware: Disaster Research Center.
6. Quarantelli, E.L (1999). *Disaster related social behavior: Summary of 50 years of research findings*. University of Delaware: Disaster Research Center.