Analytic Strategies for Collaborative Intelligence Analysis

PETER SVENMARCK, MAGDALENA GRANÅSEN,
JIRI TRNKA, HELENA GRANLUND

FOI, Swedish Defence Research Agency, is a mainly assignment-funded agency under the Ministry of Defence. The core activities are research, method and technology development, as well as studies conducted in the interests of Swedish defence and the safety and security of society. The organisation employs approximately 1000 personnel of whom about 800 are scientists. This makes FOI Sweden's largest research institute. FOI gives its customers access to leading-edge expertise in a large number of fields such as security policy studies, defence and security related analyses, the assessment of various types of threat, systems for control and management of crises, protection against and management of hazardous substances, IT security and the potential offered by new sensors.

FOI
Defence Research Agency
Information Systems
Box 1165
SE-581 11 Linköping

Phone: +46 13 37 80 00
Fax: +46 13 37 81 00

www.foi.se

FOI-R--3177--SE Scientific Report
ISSN 1650-1942 January 2011

Information Systems

Peter Svenmarck, Magdalena Granåsen,
Jiri Trnka, Helena Granlund

Analytic Strategies for Collaborative Intelligence Analysis

Titel	Analytiska strategier för gemensam analys och syntes
Title	Analytic Strategies for Collaborative Intelligence Analysis
Rapportnr/Report no	FOI-R--3177--SE
Rapporttyp Report Type	Vetenskaplig rapport/Scientific Report
Sidor/Pages	31 p
Månad/Month	Januari/January
Utgivningsår/Year	2011
ISSN	ISSN 1650-1942
Kund/Customer	Försvarsmakten
Projektnr/Project no	E53049
Godkänd av/Approved by	Magnus Jändel

FOI, Totalförsvarets Forskningsinstitut
Avdelningen för Informationssystem
Box 1165
581 11 Linköping

FOI, Swedish Defence Research Agency
Information Systems
Box 1165
SE-581 11 Linköping

Sammanfattning

Vid internationella insatser så blir det allt viktigare med tillförlitliga underrättelser och god situationsförståelse av socio-kulturella faktorer, inflytelserika aktörers motiv och intentioner och geopolitisk bakgrund. För att hantera den här komplexiteten så skapas ofta sammansatta heterogena team av specialister med olika tematiska perspektiv. Problemet är att de heterogena teamen i likhet med många enskilda analytiker sällan använder någon gemensam analytisk strategi för att genomföra analysen. Det här gör att teamen blir sårbara för vanliga problem vid gemensamma analyser som att bara ta hänsyn till information som stödjer den nuvarande uppfattningen, otillräcklig analys av alternativa möjligheter och otillräcklig granskning av väsentliga antaganden.

Den här rapporten beskriver ett experiment för att utvärdera hur de två analytiska strategierna Kritiskt tänkande och Analys av motstridiga hypoteser förbättrar den gemensamma analysen. Kritiskt tänkande uppmuntrar teammedlemmarna att med frågor och svarsdialoger förstå styrkor och svagheter med alternativa mentala modeller som integrerar observationer, tolkningar, möjliga åtgärder och förväntat resultat. Analys av motstridiga hypoteser i gengäld uppmuntrar teammedlemmarna att för varje informationsuppgift, antagande och argument utvärdera hur de stödjer eller motsäger varje hypotes. Syftet med analysen är att identifiera den hypotes eller hypoteser som motsägs minst eller inte alls.

Resultaten av experimentet visar att team som använder Kritiskt tänkande presterar bättre än team som använder Analys av motstridiga hypoteser. Team som använder Analys av motstridiga hypoteser presterade till och med sämre än team som inte använde någon explicit strategi. Det viktigaste skälet till att team som använde Analys av motstridiga hypoteser presterade sämre var att de hade svårt att bedöma komplex och interrelaterad information relativt alla potentiella hypoteser. Istället så fastnar teamen i förvirrade diskussioner och oenighet. Men även om team som använder Kritiskt tänkande presterar bäst så är deras kommunikationsmönster likartat med team som inte använder någon explicit strategi. Dessutom brister kvaliteten på många av deras hypoteser. Det finns därför all anledning att utveckla bättre strategier för gemensamma analyser.

Nyckelord: Gemensamma analyser, analytiska strategier, Kritiskt tänkande, Analys av motstridiga hypoteser, sammansatta heterogena team

Summary

Intelligence analysis is increasingly important in international crisis management which requires a considerable understanding of socio-cultural dynamics, influential actors' motives and intentions, and geopolitical realities. Due to this complexity, the intelligence analysis is often performed by heterogeneous teams of complementary specialists which may come from different domains and organizations. However, similar to individual analysts, the heterogeneous teams commonly do not use any specific analytic strategy for how to perform the collaborative intelligence analysis. The lack of analytic strategy means that the teams are vulnerable to typical problems in collaborative analysis, such as only considering information that support the current hypothesis, insufficient consideration of alternative interpretations, and incomplete assessment of key assumptions.

This report describes an experiment to evaluate how the analytic strategies Critical Thinking and Analysis of Competing Hypotheses (ACH) improve the performance in collaborative intelligence analysis. Critical Thinking encourages team members to use question and answer dialogues to understand the strengths and weaknesses of alternative mental models of the situation that integrate observations, interpretations, possible actions, and expected results. Analysis of Competing Hypotheses, on the other hand, encourages team members to evaluate each type of information, assumption, and argument relative all hypotheses to clearly identify the hypothesis with the least weaknesses.

The result of the experiment shows that teams that use Critical Thinking perform better than teams that use Analysis of Competing Hypotheses. Teams that use Analysis of Competing Hypotheses also perform worse than teams that did not use any explicit strategy. The main reason that teams that use Analysis of Competing Hypotheses perform worse is that they have difficulty in fully assessing complex and interrelated information relative all potential hypotheses. The assessment difficulties result in confusing discussions and disagreements. However, although the teams' performance improves when using Critical Thinking, their communication pattern is similar to the teams that did not use any explicit strategy and many of their hypotheses have an unsatisfactory quality. This indicates that there is still considerable potential for improving the collaborative intelligence analysis with better analytic strategies.

Keywords: Collaborative Intelligence Analysis, Analytic Strategies, Critical Thinking, Analysis of Competing Hypotheses, Heterogeneous Teams

Table of contents

1 Introduction

Intelligence analysis is increasingly important in international crisis management. Typically, such peacekeeping and stability operations, as well as aid and disaster response activities, primarily occur in countries and regions with a weakly developed government and infrastructure. Conducting operations in such complex environments require a considerable understanding of socio-cultural dynamics, influential actors' motives and intentions, and geopolitical realities. Comprehensive intelligence analysis covering political, military, economical and humanitarian aspects is therefore essential in order to support commanders and executives at different decision-levels of the operations.

Due to the complex situation in international operations, the intelligence analysis is often performed by heterogeneous teams of complementary specialists which may come from different domains and organizations. In the future intelligence analyses in the Swedish Armed Forces, such heterogeneous teams are formed by analysts with different thematic perspectives (Swedish Armed Forces, 2008). Forming heterogeneous teams improves performance since interdependent work teams in intelligence analysis outperform the more common coacting groups by 25% (Hackman & O'Connor, 2005). Additionally, heterogeneous teams can also utilize the team members' diverse experience for a more comprehensive analysis. However, similar to individual analysts, the heterogeneous teams commonly do not use any specific analytic strategy for how to perform the collaborative intelligence analysis. The lack of analytic strategy means that the teams are left with their own judgments for how to cope with the typical characteristics of intelligence analysis where there is (1) a high metal workload, (2) risk for severe consequences of a mistake, (3) time pressure, (4) a large amount of data from diverse sources, (5) non-verified and incomplete information, (6) strong context influence, and (7) complicated judgements (Hutchins, Pirolli, & Card, 2006). Additionally, different prior knowledge, access to information, and organizational and cultural background of individual team members, influence the analysis and teamwork. Neither is it unusual that teams in international operations have to work with situations and problems that are outside the teams' main field of competence (Trnka et al., 2009).

All these characteristics of collaborative intelligence analysis means that the lack of analytic strategy makes the teams vulnerable to confirmation bias of only considering information that support the current hypothesis (Lehner, Adelman, Cheikes, & Brown, 2008), Collective Information Sampling bias (CIS) towards only discussing information that is shared among the team members (Stasser & Titus, 1985), premature closure with insufficient consideration of alternative interpretations (Okhuysen & Eisenhardt, 2002; Elm et al., 2005), and incomplete assessment of key assumptions (Heuer, 2007; Trent, Patterson, & Woods, 2007). These vulnerabilities may occasionally affect the reliability of the collaborative intelligence analysis.

1.1 Objectives

Given the potential vulnerabilities of collaborative intelligence analysis, the objectives of the research presented in this report are to:

1) Identify analytic strategies that are feasible for collaborative intelligence analysis.
2) Investigate whether these strategies change the dialogue of the collaborative analysis in a way that improves the quality of the analysis.

The first objective was addressed by performing a literature review of the characteristics of collaborative intelligence analysis and potential analytic strategies for performing a collaborative intelligence analysis. The next section describes the result of this literature review. Based on the literature review two analytic strategies were selected for further evaluation, Critical Thinking and Analysis of Competing Hypotheses. The evaluation of these strategies for performing a typical collaborative intelligence analysis is described in the following section. Finally, there is a discussion and recommendations for future studies.

2 State of the art in collaborative intelligence analysis

Although there are many reasons for performing an intelligence analysis, it often serves as a sense-making process where "…individuals (or organizations) create an understanding so that they can act in a principled and informed manner" (PARC, 2009; Weick, 1995). This process may be described as a series of loops for finding and extracting information, as well as developing a mental model that fits available evidence (Pirolli & Card, 2005). The loops are often used in an opportunistic mix of bottom-up processing starting with information foraging, and top-down processing starting with sense-making. From a functional perspective, the loops are a balance of narrowing functions for identifying hypotheses that explains the interpretations of relevant data, and broadening functions for checking the accuracy of the narrowing (Elm et al., 2005).

Performing the sense-making with collaborative intelligence analysis means that the team's analytic strategy should consider several scientific disciplines, such as:

1) Known analytic strategies for specific aspects of the intelligence analysis. For example, to consider alternative possibilities, systematic evaluation of information from several sources, and challenge established views (CIA, 2009). The most well known analytic strategy is perhaps Analysis of Competing Hypotheses (ACH) (Heuer, 1999). ACH is inspired by how hypothesis testing is used within scientific research to cope with imperfect information management. ACH encourages the analyst/team to separate the generation of information, assumptions, and arguments from the actual testing of hypotheses. Contradictory information is therefore more likely to be evaluated correctly. For example, Lehner et al. (2008) report how ACH reduce the confirmation bias for participants without experience in intelligence analysis.

2) Team research regarding team decision making (Rasmussen, Brehmer, & Leplat, 1990; Orasanu & Salas, 1993), and group communication theory (Hirokawa & Poole, 1996; Frey, Gouran, & Poole, 1999). Much of the recent research is summarized within macrocognition, which is the externalized cognitive processes that the team members use to exchange and integrate knowledge, create a shared understanding, and converge their individual mental models into a team mental model (Warner & Letsky, 2008; Fiore et al., 2010).

3) Group psychology research about biases, such as the collective information sampling bias; why they occur, potential negative effects on the quality of analysis, and how to minimize the problems (e.g. Straus et al., 2009). Potential mitigations may be to modify the group structure by promoting heterogeneous groups and divergent views, use a structured procedure, facilitators, or better training (Straus et al., 2009).

4) Strategies for critical thinking that encourage reflection of the analytic process and quality of analysis (Moore, 2007; Elder & Paul, 2007). Training in experiential based critical thinking is increasingly provided to intelligence analysts and their managers (Moore, 2007; Fischer, Spiker, Harris, McPeters, & Riedel, 2008).

5) Research in group dynamics that encourages team members to express divergent views in order to identify and assess essential assumptions and uncertainties. This is the main benefit of heterogeneous teams since divergent views occur naturally, but the teams also need training in how to properly communicate and challenge each other. Creating opportunities for divergent views broadens the discussion (Sniezek & Paese, 1989), reduce the risk for Confirmation Bias (Schultz-Hardt, Jochims, & Frey 2002), improves the sharing of information which is the main

problem in Collective Information Sampling bias (Henry, 1995), and improves the quality of decisions (Amason & Schweiger, 1997; Jehn, 1997).

One collaborative analytic strategy that covers many of these aspects is Critical Thinking based on dialogue theory as proposed by Cohen et al. (2007). The advantage of this analytic strategy is that it combines two pertinent trends for improving the reliability of the intelligence analysis: the use of interdependent work teams and stimulation of critical thinking (Svenmarck, 2011). The main benefit of using dialogue theory is that it formalizes the team members' motives, as well as the structure, constraints, and rules for typical dialogues (Walton, 1998; van Eemeren & Grootendorst, 1992). Initially, the team members develop mental models of the situation that integrate observations, interpretations, possible actions, and expected results. For critical disagreements between the mental models, a reasoned dialogue is then performed between a proponent and opponent while monitored by a facilitator who assesses the relevance of arguments and rebuttals and guides the dialogue towards the overall goals of the analysis. Cohen et al. report how teams that receive training in critical thinking are more likely to recognize and deal with disagreements, to ask for and give reasons for positions, and seek creative solutions rather than premature compromises.

Clearly, Critical Thinking based on dialogue theory is a potentially useful analytic strategy for collaborative intelligence analysis. However, although Critical Thinking may be particularly suitable for collaborative intelligence analysis by heterogeneous teams, there are also several other analytic strategies that may be useful. For example, an overview of analytic strategies may be found in Jones (1998), Snowden, Klein, Pin, & Teh (2007), and CIA (2009). One of the best known of these analytic strategies is Analysis by Competing Hypotheses. Since no studies have compared these analytic strategies, an experiment was performed to evaluate the effects of using Critical Thinking, Analysis of Competing Hypotheses, and not using an explicit strategy when teams are confronted with a typical collaborative intelligence analysis problem. The experiment to evaluate the analytic strategies is described next.

3 Evaluation of selected analytic strategies

The literature review shows that Critical Thinking and Analysis of Competing Hypotheses (ACH) are promising analytic strategies for collaborative intelligence analysis. Previous studies have also shown positive effects on the problem analysis when using Critical Thinking (Cohen et al., 2007) and Analysis by Competing Hypotheses (ACH) (Lehner et al., 2008). An experiment was therefore performed to compare the effects of these strategies on collaborative intelligence analysis. The hypotheses of the experiment were that:

1) Teams that use the analytic strategies Critical Thinking or Analysis by Competing Hypotheses perform better than teams that do not use any explicit analytic strategy.
2) There is only a small difference in performance between teams that use Critical Thinking and teams that use Analysis by Competing Hypotheses, since both analytic strategies address a similar problem.

The hypotheses were evaluated by comparing how teams with and without training in the analytic strategies perform a typical collaborative intelligence analysis problem. Although many types of scenarios and problems have been used in studies of intelligence analysis, the main requirements for the present experiment was that the scenario should not require domain expertise nor experience in intelligence analysis and still provide a challenging problem with uncertain and incomplete information that needs to interpreted and assessed in order to gain an overall understanding. One suitable scenario based on these requirements is the investigation of the cause or causes of a mysterious disease among the native population on Guam (Suthers et al., 2007). The scenario material was used with permission from Prof. Daniel Suthers.

3.1 Experiment design

The experiment evaluated the analytic strategies Critical Thinking or Analysis of Competing Hypotheses in comparison with a control group where the participants did not use any explicit analytic strategy. Since the teams only analyzed one problem, the analytic strategy (Critical Thinking, ACH, or None) was the only independent measure in a between-group design. The only difference between the three conditions was the analytic strategy that the participants used or that they did not use any explicit analytic strategy.

3.1.1 Participants

The teams worked in groups of three with five teams in each condition. A total of 45 university students participated in the experiment. Totally, 30 women and 15 men participated in the experiment. There were six all female teams, two all male teams, and seven mixed teams.

3.1.2 Procedure

Figure 1 shows the experimental procedure that consisted of Introduction, Training, Experiment Scenario and Closure. The experiment as a whole took approximately four hours.

During the *Introduction*, the participants received information about the purpose of the experiment and the experimental procedure in its entirety. They also filled in two questionnaires, a pre-experiment questionnaire and a personality test using NEO FFI. The introduction took about 30 minutes.

After the introduction, the training was performed in three steps. First, the participants received a 20-minute lecture about one of the two analytic strategies, as well as how to use the strategy. Then, the participants performed two training scenarios in order to gain practical experience of using the analytic strategy. The first training scenario concerned route planning in the face of uncertainty. During the first training scenario, the lecturer and co-observer guided the participants about how they were supposed to use the analytical strategy. The first training scenario lasted for about 20 minutes. The second training scenario concerned an aircraft accident investigation and lasted for about 30 minutes. During the second training scenario, the participants used the strategy without guidance or instruction, but were free to ask questions to the lecturer. The participants had approximately ten minutes for joint reflection after the lecture and after every training scenario. The teams in the control group did not receive any lecture. Instead, they played billiards and had team-talk during the same time period. Otherwise, the teams in the control group performed the same training scenarios. Totally, the *Training* took about two hours.

After the Training, the participants performed the *Experiment Scenario*. The experiment scenario started with that the lecturer explained the participants' task where they had to analyze the situation and then identify and justify plausible causes for the health problems on Guam (see Suthers et al., 2007). The participants were also asked to collectively rank the identified causes according to their plausibility. Each participant received a unique folder with individual information about the situation. The participants therefore had to share their information in order to jointly solve the task. The participants worked together and were free to solve the task on their own when the scenario had started. The teams recorded their final hypotheses for the cause, the rank order, and justifications on a laptop. The Experiment Scenario lasted for about 80 minutes.

The last activity was *Closure* where the participants filled in a post-experiment questionnaire. Then, they were provided a short general feedback regarding their participation in the experiment and given an opportunity for questions.

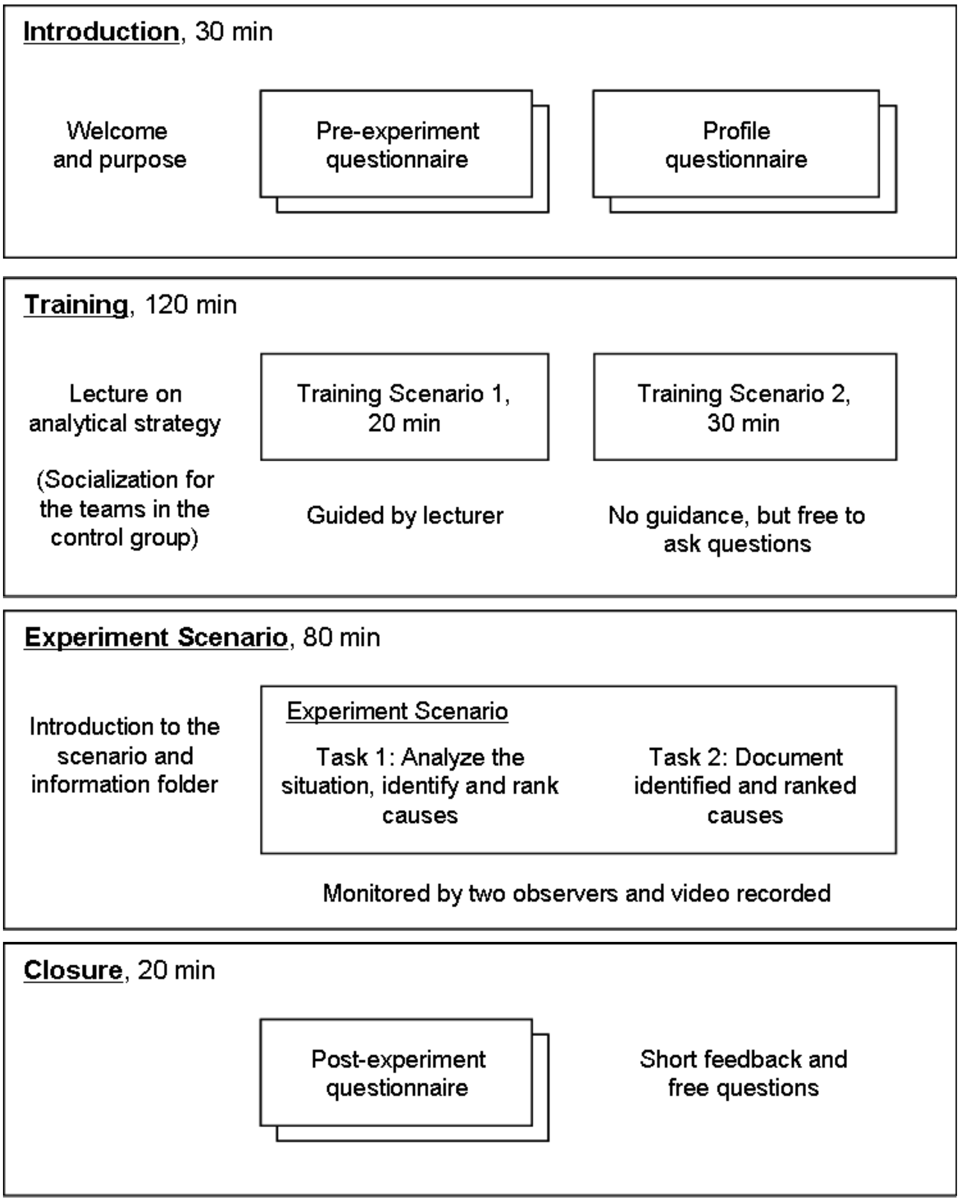

Figure 1 The experiment procedure.

3.1.2.1 Lecture on analytic strategy

The training in using the analytic strategy started with a 20-minute lecture about one of the two analytic strategies, Critical Thinking or Analysis of Competing Hypotheses. The lecture included a brief theoretical review of the respective strategies' principles and process.

The reasons for using Critical Thinking as a strategy for collaborative intelligence analyses are to avoid conventional solutions and habits from dominating the discussion without questioning, to encourage a structured analysis, and to provide a better understanding of the solutions' strengths and weaknesses. The principles of Critical Thinking in collaborative intelligence analyses is to use question and answer dialogues to refine the understanding of alternative possibilities while achieving the objectives of the analysis. Critical Thinking as a process consists of four stages: (1) identify the problem area, (2) prioritise areas for critical dialogue, (3) challenge and defend, and finally (4) make a decision (Cohen et al., 2007). The core of Critical Thinking is the challenge and defend phase where diverging views are contrasted using recommended principles for an efficient critical dialogue. The diverging views are assessed both in terms how they comply with intellectual standards such as clarity, accuracy, logic, and significance (Paul & Elder, 2008), their representation of the uncertainty when accounting for the available information, as well as their overall plausibility (Cohen et al., 2007).

The reasons for using Analysis of Competing Hypotheses are to enable a systematic assessment of all the information, avoid premature hypotheses selection, and to provide a structured justification for the recommended solution. The principles of ACH are that the hypotheses are formulated from existing knowledge, evaluated on the basis of available evidence, and falsified rather than confirmed. The process consists of eight steps: (1) identification of hypotheses, (2) list all the important arguments, (3) evaluate each argument against all hypotheses, (4) refine the table by clarifying hypotheses, removing arguments that do not contribute to the hypotheses evaluation, and adding additional arguments that do contribute to the hypotheses evaluation, (5) compare hypotheses, (6) review the results, (7) identify indicators for monitoring, and (8) report the results (Heuer, 1999, cf. Jones, 1998). The core of ACH is the separate analysis of each argument relative all hypotheses to clearly identify the hypothesis with the least weaknesses.

3.1.2.2 Training scenarios

The teams analysed three scenarios during the experiment, where two scenarios were used for training one for the actual experiment. The scenarios had an increasing level of complexity, where the experiment scenario was the most complicated. Overall, the scenarios are typical of real world problems where there is incomplete, uncertain and conflicting information that require analysis at several levels. For each scenario, the participants received individual and unique information about the scenario, jointly analysed the problem, identified possible options with justifications, and rank ordered the options. All scenarios also had a map showing the overall situation. The participants were free to take notes, underline important information, use sticky notes, and a whiteboard for discussions. The teams that received training in ACH also had a matrix that they could use to structure the information about hypotheses and argument as recommended by the strategy.

The first training scenario was set in the fictional region "Palombi" where the participants were responsible for a peacekeeping patrol of 10 people. The patrol's mission is to get from its current location to the planned destination using an alternative route since the originally planned route is damaged by flooding. The patrol is faced with the choice of one of three possible routes that each have their advantages and disadvantages regarding time of arrival, risk of attack, and tactical possibilities. The material was presented in Swedish.

The second training scenario described an aircraft accident outside New York in 1996. The participants were responsible identifying possible causes of the accident and rank the causes according to their plausibility. There are several conflicting sources of information that indicate structural failures, terrorist attack, and human error. The evidence is not conclusive for any specific cause, but allows an assessment of their plausibility. The material was presented in English.

3.1.2.3 Experiment scenario

The final experiment scenario concerned an investigation of a real problem regarding the cause or causes of a mysterious disease among the native population on Guam (Suthers et al., 2007). The information indicates various disease agents and vectors of introduction. While some hypotheses can be ruled out, others are more or less plausible. No definite explanation is available to this day, but one is more plausible than the others. The scenario material was adapted to the constraints of the present experiment, primarily by reducing the amount of information by removing all information about the disease agent aluminium. The material was presented in English.

3.1.3 Data collection

The data collection during the experiments consisted of three participant questionnaires (pre-experiment, personality test, and post-experiment), two observer questionnaire (lecturer and co-observer post-experiment), observation and categorization of communication and problem analysis, audio and video recordings, and archiving the participants notes.

The pre-experiment questionnaire was a demographic questionnaire with 9 items. The questions concerned the participants' personal background and experiences of analytic work, particularly regarding the two strategies, *Critical Thinking* and *ACH*, that were evaluated in the experiment. The personality test was the 60 item NEO Five Factor Inventory (NEO-FFI) (Costa & McCrae, 1992). The NEO-FFI measures five areas of adult personality: extraversion, agreeableness, conscientiousness, emotional stability, and intellect/openness. The participants post-experiment questionnaire was divided into teamwork (19 items), use of the analytic strategy (10 items), and the adequacy of the training in the analytic strategy (1 item). Most of the items were rated on five point Lickert-scales. Items regarding the contribution and domination of team members were rated using percentages. Similarly, items regarding the analytic process were rated using percentages.

During the Experiment Scenario two independent observers (the lecturer and co-observer) simultaneously monitored the team's actions. The lecturer monitored and categorized the team's communication while the co-observer recorded more general events of the team's problem analysis. Both observers recorded their observations of the experiment on a laptop using the Network-Based-Observer-Tool (NBOT) (Thorstensson, 2008)). NBOT is designed for recording sequential observations in real time during an experiment. Two NBOT schemas were configured specifically for the experiment and were installed on the observers' laptops. NBOT schema for the communication covered collaboration (4 items), hypotheses (3 items), decision-making (3 items), discussion, and communication that was not covered by the previous items. The NBOT schema for the general events covered the team’s teamwork (4 items), use of strategy, discussion, and events that were not covered by the previous items. The co-observer also recorded the time for completing the activities during the experiment, as well as the time for completing the steps of the analytic strategy. These steps were emphasized during the training of the analytic strategy or identified by the co-observer for the teams in the control group that did not use any explicit analytic strategy. The experiment was also audio and video recorded as shown in Figure 2.

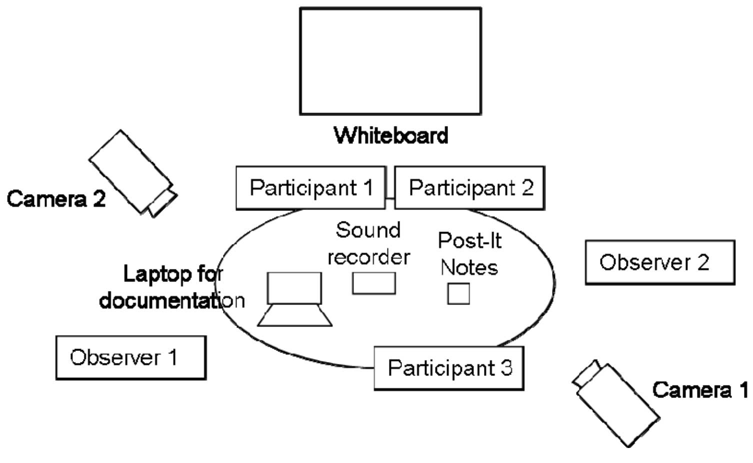

Figure 2 Overview of the apparatus, participants, and observers.

The observers' post-experiment questionnaire assessed the team performance in three areas: teamwork (21 items), use of the analytical strategy (3 items), and discussion of hypotheses (3 items). Most of the items were rated on five point Lickert-scales. Items regarding the contribution and domination of team members were rated using percentages.

Similarly, items regarding the analytic process were rated using percentages. Finally, items regarding whether the teams missed or misinterpreted information and identified the relevant hypotheses were rated using yes or no.

3.1.4 Dependent measures

The most plausible disease agent for the ALS-PD disease among native Guamians is a neurotoxic amino acid in the seed of the Cycad tree. The most plausible vector for the introduction of the agent is consumption of bats that eat the seeds. The teams' performance was therefore measured by their ranking of a hypothesis that considers both the most plausible disease agent and vector. The performance was measured by whether the most plausible hypothesis received the highest ranking, as well as the ranking for the most plausible hypothesis.

Most items in the participants' questionnaires were normalized by computing the average of the team members' ratings. Similarly, items in the observers questionnaires were normalized by the computing the average of the observers' ratings. Further, for some items the participants rated each of the individual team members. The computation of the team measure for these items was therefore preceded by computing the average of each participant's rating of the team members. This was the case for the pre-experiment questionnaires items regarding previous experience of working together and familiarity, the participants' post-experiment questionnaire regarding the involvement of team members, and the observers' post-experiment questionnaire regarding how the team members followed each others suggestions, involvement, and how comfortable the team members were with the final decision. Finally, for questions regarding the contribution and domination of team members, the computation of the team measure was preceded by computing the standard deviation of each participant's rating to capture the perceived differences between team members.

3.1.5 Methods of analysis

An ANOVA was performed for most of the dependent measures. The ANOVA shows whether there is a significant difference or differences between the strategies, but not between which strategies there is a significant difference. Therefore, a Tukey HSD test was performed for the significant ANOVA to identify between which strategies there were significant differences. The probability for incorrectly identify a significant difference between the strategies was 0.05 or lower as indicated in the results. However, since only five teams participated in each condition, only the major differences were statistically significant. Potentially interesting smaller differences were therefore tested by whether the standard error bars did not overlap. The only exception to the ANOVA was for whether the most plausible hypothesis received the highest ranking, where a Cochran Q Test was used. The Cochran Q Test is a non-parametric statistical test of whether outcomes that either occur or do not occur are statistically significant between the conditions. Finally, no analysis was performed of the general events that were registered by the co-observer, nor of the audio and video recordings.

3.2 Results

3.2.1 Pre-experiment inquiry

The results from the pre-experiment questionnaire show that participants were between 18 and 33 years of age with an average of 24.4 years. 53% of the participants reported that they had previous experience of analytic work, primarily during their studies. Two participants reported that they had experience of Critical Thinking, but only one of these had received a formal training.

3.2.1.1 Team familiarity

Table 1 shows the average and ANOVA for the pre-experiment familiarity. The table shows that the participants in the teams that use ACH knew each other more than the participants in the other conditions. A Tukey HSD test shows that the teams that use ACH knew each other significantly more than the teams that did not use any explicit strategy ($p<0.05$). The ACH teams' average previous experience of working together was 2.23, which is between *Rarely of once or twice a year* and *Sometimes 10-15 times/year* on a five point Likert-scale ranging from *Never* to *Constantly every day*. The highest average previous experience of working together for a team was 3.0, which corresponds to *Sometimes 10-15 times/year*, and the lowest was 1.0, which corresponds to *Never*. Since the participants in the ACH teams have more experience of working together it is not surprising that they also report a higher level of acquaintance. The average whether the ACH participants knew each other before the experiment was 2.93, which is between *Recognize* and *Met at some occasion* on a five point Likert-scale ranging between *Never met* before and *Know very well*. The highest average rating for whether the team members knew each other was 4.33, which is between *Know* and *Know very well*, and the lowest was 1.0, which corresponds to *Never met*. Generally, however, these differences in familiarity did not affect how comfortable the participants were working together.

Table 1 Average and ANOVA for pre-experiment familiarity. * indicate non-overlapping ± standard error bars. ** indicate significant at p<0.05. ns indicate not significant.

Familiarity	**Critical Thinking**	**ACH**	**No Explicit**	**F(2, 12)**	**p**
Worked together previously	1.47	**2.23**	1.07	6.26	0.014**
Know each other	2.20	2.93	1.63	1.67*	0.230
Comfortable working together	4.00	4.20	4.00	ns	

3.2.1.2 Personality test

Table 2 shows the result from the NEO FFI personality test. The table shows that the participants in teams that use Critical Thinking rate themselves as slightly less open than the participants in the other teams. Further, the participants in teams that did not use any explicit strategy also rate themselves as slightly more agreeable and conscientious than the participants in the other teams.

Table 2 Results from the NEO FFI personality test. * indicate non-overlapping ± standard error bars. ns indicate not significant.

NEO FFI Dimension	**Critical Thinking**	**ACH**	**No Explicit**	**F(2, 12)**	**p**
Neuroticism	14.5	19.3	18.4	ns	
Extraversion	33.0	35.4	31.6	ns	
Openness	27.8	30.6	30.8	2.19*	0.154
Agreeableness	35.9	35.4	38.4	1.64*	0.235
Conscientiousness	33.3	32.7	35.4	1.58*	0.246

Table 3 shows the correlation between the average previous experience of working together, average whether the participants knew each other, and the NEO FFI dimensions. The only significant correlations are between previous experiences of working together, whether the participants known each other, and extraversion. Presumably, the students mostly meet during their studies and extraversion facilitates contact between the students.

Table 3 Correlation between the average previous experience of working together, average whether the participants knew each other, and the NEO FFI dimensions. * indicates a significant correlation with p<0.05.

	Experience	Know each other	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Experience	1.00	0.79*	-0.04	0.66*	-0.22	-0.15	0.12
Know each other	0.79*	1.00	-0.14	0.70*	-0.18	0.04	0.33
Neuroticism	-0.04	-0.14	1.00	-0.44	0.02	-0.14	-0.16
Extraversion	0.66*	0.70*	-0.44	1.00	0.12	0.22	0.45
Openness	-0.22	-0.18	0.02	0.12	1.00	-0.16	0.21
Agreeableness	-0.15	0.04	-0.14	0.22	-0.16	1.00	0.12
Conscientiousness	0.12	0.33	-0.16	0.45	0.21	0.12	1.00

3.2.2 Quality of analysis: Identified and ranked causes

Table 4 shows the percentage of teams with the highest ranking for the most plausible hypothesis and the average ranking for the most plausible hypothesis. The table shows that teams that use Critical Thinking perform better than both teams that use ACH and teams that did not use any explicit strategy. Surprisingly, teams that use ACH even perform worse than teams that did not use any explicit. However, all teams perform better than in the original study by Suthers et al. (2007) where on average only 15% gave the highest ranking to the most plausible hypothesis. This is most likely due to that the participants had less information to consider in the present study since all information about aluminium was removed and there were three participants in each team where as there were only two participants in each team in the study by Suthers et al. A Cochran Q Test of the highest ranking for most plausible hypothesis did not indicate any significant difference ($Q = 2.00$, $df = 2$, $p < 0.368$). Neither was an ANOVA significant of the ranking for the most plausible hypothesis ($F(2, 12)=1.41$, $p=0.282$). However, the non-overlapping standard error bars in Figure 3 between Critical Thinking and ACH teams indicate that the difference may be significant if more teams performed the experiment.

Table 4 Percentage with the highest ranking for the most plausible hypothesis and average ranking for the most plausible hypothesis (lower is better).

Performance Measure	Critical Thinking	ACH	No Explicit
Percentage with highest ranking for most plausible hypothesis	80%	40%	60%
Average ranking for most plausible hypothesis	1.2	2.2	1.6

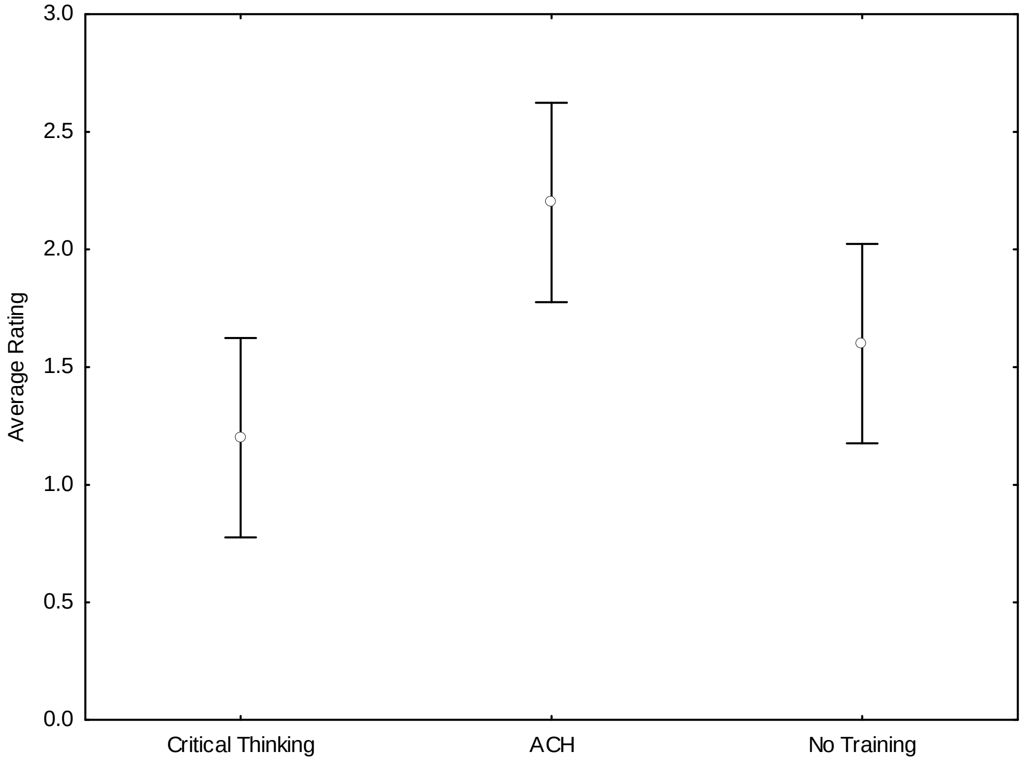

Figure 3 Average ranking for most plausible hypothesis. The bars denote ± standard error bars.

Table 5 shows the correlation between the ranking for the most plausible hypothesis, average previous experience of working together, average whether the participants knew each other, and the NEO FFI dimensions. The correlation between the ranking for the most plausible hypothesis and average previous experience of working together was 0.12. Similarly, the correlation between the ranking for the most plausible hypothesis the average whether the participants knew each other was 0.18. Although familiarity contributes to performance, other factors are clearly more important. Further, neither was there any significant correlation between the ranking for the most plausible hypothesis and the NEO FFI dimensions. Finally, Table 6 shows the result of a multiple regression analysis for whether the ranking for the most plausible hypothesis can be predicted from the familiarity and NEO FFI dimensions. The multiple regression analysis was not significant ($F(7,7)=1.46$, $p<0.32$) although the explained variance was reasonable ($R^2= 0.59$). The only predictors with modest influence were openness and neuroticism.

Table 5 Correlation between the ranking for the most plausible hypothesis, average previous experience of working together, average whether the participants knew each other, and the NEO FFI dimensions.

	Ranking
Ranking	1.00
Experience	0.12
Know each other	0.18
Neuroticism	-0.26
Extraversion	0.11
Openness	0.47
Agreeableness	-0.40
Conscientiousness	0.16

Table 6 Result of a multiple regression analysis for whether the ranking for the most plausible hypothesis can be predicted from the familiarity and NEO FFI dimensions.

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(7)	p-level
Intercept			-0.03	5.65	-0.01	1.00
Experience	0.38	0.56	0.53	0.78	0.68	0.52
Know each other	0.47	0.45	0.39	0.37	1.05	0.33
Neuroticism	-0.55	0.32	-0.08	0.05	-1.74	0.13
Extraversion	-0.82	0.60	-0.17	0.13	-1.36	0.21
Openness	0.71	0.32	0.25	0.11	2.19	0.06
Agreeableness	-0.15	0.31	-0.05	0.11	-0.48	0.65
Conscientiousness	0.11	0.29	0.04	0.11	0.38	0.71

Some clues about why some teams did no give the highest ranking to the most plausible hypothesis may be gained from an inspection of the teams' written reports where they described their arguments for and against the hypotheses. The inspection shows that all teams performed fairly well in assimilating the facts about potential disease agents and vectors of introduction. However, the best way to distinguish among all these hypotheses is to evaluate them against the disease pattern. This disease pattern covers the populations or sub-populations that are affected by the disease, under what circumstances, demographic differences, and how the disease pattern changes over time. Facts and reasonable assumptions about the hypotheses should preferably explain the disease pattern or at least not contradict with the disease pattern. Inspection of the teams' reports shows that the main reason for the reduced quality of analysis for some teams was a missing, irrelevant, limited, or erroneous understanding of the disease pattern. The teams' actual evaluation of the hypotheses based on their understanding of the disease pattern was, in turn, usually satisfactory. Any further understanding of why particularly the ACH-teams have difficulty assimilating information about the disease pattern can only be gained from an analysis of the teams' communication, which is beyond the scope of the current report. Such an analysis may also explain the analytic errors for the successful teams regarding the less plausible hypotheses. Here, the successful teams often have an erroneous understanding of the facts, make questionable assumptions, or have an insufficient consideration of the disease pattern. There is therefore a considerable potential for improving the communication even among the successful teams.

3.2.3 Experiment observations

3.2.3.1 Time for experiment activities

Table 7 shows the average time for the phases during the experiment. The only significant difference is that training takes about 10-15 minutes longer for teams that receive training in ACH since the strategy contains more steps. A Tukey HSD test shows that the training for the teams that use ACH takes significantly longer than for the teams that did not use any explicit strategy ($p<0.05$).

Table 7 Average time for the phases during the experiment. ** indicate significant at p<0.05. ns indicate not significant.

Phase	Critical Thinking	ACH	No Explicit	F(2, 12)	p
Introduction	00:21	00:22	00:23	ns	
Training	01:53	**02:01**	01:46	5.45	0.021**
Experiment Scenario	01:26	01:23	01:23	ns	
Closure	00:18	00:21	00:18	ns	

3.2.3.2 Use of analytic strategies

Table 8 to Table 10 shows the observer's indicated time for completing the analytic steps. The tables show that the teams generally spend most of their time on information assimilation and problem structuring. Most of the hypotheses evaluation is embedded in these discussions and is only performed explicitly for a short time. Teams that use Critical Thinking only use 15% of the time to challenge and defend hypotheses. Teams that use ACH only use 10% of the time to compare hypotheses and teams that did not use any explicit strategy only use 20% of the time for arguments for and against. Further, the strategy steps are largely problem driven for the teams that did not use any explicit strategy. The reason that the sum of the time for steps is larger than the total time is that the steps are overlapping since it often was difficult to identity when the teams went from one step to another.

Table 8 Average time for the Critical Thinking strategy steps.

Simulation Step	Average Time
Information assimilation	00:24
Identify problem areas	00:22
Prioritize areas for critical discussion	00:11
Challenge-defend	00:12
Decision	00:16
Sum	01:28
Total Time	01:19

Table 9 Average time for the ACH strategy steps.

Simulation Step	Average Time
Information assimilation	00:23
Identify hypotheses	00:24
List significant arguments	00:18
Create table	00:19
Refine table	00:10
Compare hypotheses	00:08
Review conclusions	00:10
Report conclusions	00:07
Sum	02:00
Total Time	01:21

Table 10 Percentage of teams that used the strategy steps and the average time for the strategy steps that were identified by the observer in teams that did not use any explicit strategy.

Simulation Step	Percentage	Average Time
Arguments for and against	60%	00:17
Decision	100%	00:07
Discussion/argumentation	40%	00:25
Evidence structuring	20%	00:20
Hypotheses ranking	40%	00:20
Hypothesis generation	100%	00:18
Hypothesis valuation	20%	00:00
Information assimilation	100%	00:19
Information integration	60%	00:25
Information sharing	80%	00:26
Report	20%	00:05
Sense making	20%	00:05
Validation/quality control	40%	00:05
Sum	NA	03:13
Total Time	NA	01:21

3.2.3.3 Communication analysis

Table 11 and Table 12 show the average and ANOVA for the number of communication categories and the average time for the communication categories during the experiment, respectively. The tables show that the teams identify about the same number of hypotheses in all conditions, although the teams that use ACH, as instructed, identify the hypotheses earlier than the other teams. A Tukey HSD test shows that the teams that use ACH identify the hypotheses significantly earlier than the teams that Critical Thinking ($p<0.01$) and significantly earlier than the teams that did not use any explicit strategy ($p<0.01$). Table 12 also shows that teams that use an analytic strategy take initiatives throughout the experiment, which should be beneficial for the team performance.

However, there is a disturbing trend in the communication for the teams that use ACH that likely is detrimental to their performance. Table 11 and Table 12 show that they have less argumentation both for and against hypotheses and this argumentation occurs earlier in the experiment. Further, the teams that use ACH have more disagreements and these disagreements occur throughout the experiment. The combination of these effects indicates the difficulties for the teams that use ACH. Another disturbing trend is that the amount of argumentation and the average time for argumentation is similar for teams that use Critical Thinking and teams that did not use any explicit strategy. This indicates that the Critical Thinking teams may have difficulty following the strategy. If the Critical Thinking teams actually followed the strategy, the argumentation would generally occur later during the experiment. There is a slight positive effect, however, of using Critical Thinking for the focus on the problem analysis. Although the teams that use Critical Thinking have the same amount of communication that is not related to the problem as the other teams, they have this communication in the beginning of the experiment.

Table 11 Average and ANOVA for the number of communication categories during the experiment. * indicate non-overlapping ± standard error bars. ns indicate not significant.

Communication Category	Critical Thinking	ACH	No Explicit	F(2, 12)	p
Cooperation - Disagreement	0.2	**1.4**	0.2	2.25*	0.148
Cooperation - Dominate	1.8	1.4	1.2	ns	
Cooperation - Recessive	0.4	0.8	0.2	ns	
Cooperation - Takes initiative	6.0	6.6	4.6	ns	
Decision - Consensus	1.4	1.2	1.4	ns	
Discussion	9.6	9.6	13.2	ns	
Hypothesis - Argumentation AGAINST	7.8	**4.8**	7.4	ns	
Hypothesis – Argumentation FOR	12.0	**9.6**	11.4	ns	
Hypothesis - Present NEW	7.6	7.6	7.4	ns	
Other	2.2	2.6	2.4	ns	

Table 12 Average and ANOVA for the average time in minutes for the communication categories during the experiment. * indicate non-overlapping ± standard error bars. *** indicate significant at p<0.001. ns indicate not significant.

Communication Category	Critical Thinking	ACH	No Explicit	F(2, 12)	p
Cooperation - Disagreement	6.2	**30.0**	12.6	1.26*	0.318
Cooperation - Dominate	16.0	27.3	12.0	ns	
Cooperation - Recessive	4.7	20.9	8.0	ns	
Cooperation - Takes initiative	**31.3**	**31.4**	17.2	3.60*	0.059
Decision - Consensus	25.7	24.6	28.8	ns	
Discussion	24.9	28.8	29.4	ns	
Hypothesis - Argumentation AGAINST	28.9	**22.8**	27.0	ns	
Hypothesis – Argumentation FOR	28.2	**22.9**	30.7	2.15*	0.159
Hypothesis - Present NEW	16.7	**5.1**	14.0	13.2	0.001***
Other	**5.2**	29.0	19.5	2.48*	0.125

3.2.4 Post-experiment inquiry

3.2.4.1 Participants' self-assessment

Table 13 to Table 15 shows the average and ANOVA for the participants' post-experiment questionnaire that covered teamwork, use of the analytic strategy, and the adequacy of the training. The tables show some disturbing differences and trends where using ACH has detrimental effects on the teamwork. The teams that use ACH appear to have the same understanding of the problem and workload as the other teams, but fail to identify a satisfying solution. This is primarily due to a partially inappropriate organization of the information which in turn resulted in a less comprehensive analysis. Although the main purpose of the hypotheses/argument matrix was to support a better organization of the information, it appears to be ill-suited for the current problem. The higher dominance for teams that use ACH may partly be due to that one person is responsible for the joint hypotheses/argument matrix.

Further, the teamwork is mostly similar for teams that use Critical Thinking and teams that did not use any explicit strategy. Both conditions encourage team participation, organization of information and create opportunities for a comprehensive analysis of possible hypotheses and identifying a plausible hypothesis that the whole team is comfortable with. A Tukey HSD test shows that the teams that use ACH think their hypotheses are significantly less comprehensive than the teams that use Critical Thinking ($p<0.01$) and significantly less comprehensive than the teams that did not use any explicity strategy ($p<0.05$). It is not clear why the teams that did not use any explicit strategy think that the problem is more difficult than for teams that used an explicit strategy, nor while they at the same time are better at evaluating the information. There is also a trend that teams that use Critical Thinking are more accepting towards each others suggestions, perhaps as a result of a clearer argumentation. The participants' comments indicate that the training in Critical Thinking was adequate, but the complex problem required all their attention which made it difficult to apply the strategy as intended. The participants' comments for the teams that did not use any explicit strategy, on the other hand, indicate that they would like more help in structuring the information.

Finally, it is not clear whether the differences between the conditions are due to that the training in ACH is less adequate than the training in Critical Thinking or less adequate than the socialization for the teams that did not use any explicit strategy. The participants' comments regarding the training in ACH indicate that they would have appreciated more structured training and more examples. Some participants also mention that they would have appreciated a better understanding of the differences between hypotheses and arguments. The participants seem to have different interpretations of this distinction.

Table 13 Average and ANOVA for the teamwork items in the participants' post-experiment questionnaire. * indicate non-overlapping ± standard error bars. *** indicate significant at $p<0.001$. ns indicate not significant.

Teamwork Item	**Critical Thinking**	**ACH**	**No Explicit**	**F(2, 12)/ F(1, 8)**	**p**
How did the team collaboration work?	4.3	3.9	4.5	ns	
To what extent did you feel comfortable with the team's way of working?	4.1	3.9	4.3	ns	
To what extent did team members agree on how to solve the problem?	4.1	3.7	4.1	ns	
To what extent did team members presented diverging views?	3.6	3.9	3.8	ns	
To what extent did the other team members agree with your suggestions?	**3.9**	3.3	3.4	1.90*	0.192
How did the team dialogue work?	4.3	**3.9**	4.2	ns	
What was your workload?	3.7	3.7	3.7	ns	
To what extent did you have sufficient time to solve the problem?	2.1	1.9	2.2	ns	
How was the individual team member's involvement in the experiment?	4.6	**4.3**	4.7	1.25*	0.322
How was the whole team's involvement in the experiment?	4.7	4.5	4.7	ns	

To what extent did the individual team members contribute to the solution?	4.3	6.9	4.6	ns	
Describe the dominance relationship between team members	6.8	**11.3**	7.4	1.21*	0.333
How did you evaluate the information?	3.9	3.8	**4.3**	1.26*	0.319
How did you organize the information?	4.0	**3.1**	3.9	2.02*	0.175
How do you rate your understanding of the problem?	3.9	3.7	4.0	ns	
How do you rate the difficulty of the task?	3.9	3.7	**4.2**	1.21*	0.332
How comprehensive was your analysis of the possible causes of the disease?	3.9	**2.7**	3.7	9.54	0.003***
How do you rate the team's solution?	4.1	**3.2**	4.1	2.89*	0.095
To what extent are you comfortable with the team's solution?	4.1	**3.3**	4.1	3.35*	0.070

Table 14 Average and ANOVA for the strategy items in the participants' post-experiment questionnaire. * indicate non-overlapping ± standard error bars. ns indicate not significant.

Strategy Item	**Critical Thinking**	**ACH**	**No Explicit**	**F(2, 12)/ F(1, 8)**	**p**
To what extent did you use the analytic strategy that you received training in?	3.7	3.5	NA	ns	
To what extent did you have sufficient understanding of the analytic strategy in order to utilize it the best?	3.8	3.7	NA	ns	
To what extent was the analytic strategy similar to methods that you have used before?	**3.3**	2.7	NA	2.61*	0.145
How has the use of the strategy influenced the dialogue?	**4.1**	3.7	NA	1.28	0.290
How has the use of the strategy influenced the collaboration?	3.9	3.7	NA	ns	
How has the use of the strategy, affected how you valued/assessed your information?	3.9	3.8	NA	ns	
How has the use of the strategy influenced the way you organized/ structured your information?	4.1	3.9	NA	ns	
Had the use of the strategy any effect on the number of identified and possible causes of the disease?	3.6	3.6	NA	ns	
Has the use of the strategy influenced your choice of the most plausible cause?	3.4	3.6	NA	ns	
To what extent did the analytic strategy supported the various analysis and decision-making processes?					
a. Dialogue	18.0	14.3	NA	ns	
b. Collaboration	**18.0**	14.5	NA	3.20*	0.112
c. Valuation / assessment of information	**19.0**	16.5	NA	1.58	0.244
d. Structuring / organization of information	17.0	**23.0**	NA	2.77*	0.135
e. Identification of options (reasons)	16.0	16.3	NA	ns	
f. Decision making (choice of the most plausible cause)	**12.0**	14.7	NA	1.03	0.339

Table 15 Average and ANOVA for the training item in the participants' post-experiment questionnaire. * indicate non-overlapping ± standard error bars.

Training Item	**Critical Thinking**	**ACH**	**No Explicit**	**F(2, 12)/ F(1, 8)**	**p**
To what extent was the lecture and training in the analytical approach adequate?	4.0	3.5	4.0	2.28*	0.145

3.2.4.2 Assessment by observers

Table 16 to Table 18 shows the average and ANOVA for the observers' post-experiment questionnaire that covered teamwork, use of the analytic strategy, and the completeness of identified hypotheses. Generally, the tables confirm the results from the participants' post-experiment questionnaire. Although the teams that use ACH identity as many hypotheses as the other teams, they have a worse dialogue and less agreement which contributed to their difficulty in organizing and evaluating the information and understanding the problem. These problems reduce the quality of their analysis. The observers' comments indicate that one team even completely abandoned the strategy.

Tukey HSD tests shows that the observers perceive the teams that use ACH as (1) having a significantly worse dialogue than the teams that did not use any explicit strategy (p<0.05), (2) significantly less involved that the teams that did not use any explicit strategy (p<0.05), (3) significantly worse at evaluating information than the teams that use Critical Thinking (p<0.05) and significantly worse at evaluating information than the teams that did not use any explicit strategy (p<0.01), (4) significantly worse at organizing the information than the teams that did not use any explicit strategy (p<0.05), (5) having a significantly worse understanding of the problem than the teams that use Critical Thinking (p<0.05) and significantly worse understanding of the problem than the teams that did not use any explicit strategy (p<0.05), (6) having a significantly less comprehensive analysis than the teams that did not use any explicit strategy (p<0.01), and (7) having a significantly worse solution than the teams that use Critical Thinking (p<0.05) and significantly worse solution than the teams that did not use any explicit strategy (p<0.01). Tukey HSD tests also show that the observers perceive the teams that did not use any explicit strategy as missing significantly less information than the teams that use ACH (P<0.05), and significantly undervalue important information less than the teams that use ACH (p<0.05).

Further, contrary to the participants' own ratings, the observers perceive the teams that did not use any explicit strategy as being significantly more accepting of each others suggestions. A Tukey HSD test shows that the observers perceive the teams that did not use any explicit strategy as significantly more accepting of each others suggestions that the teams that use Critical Thinking (p<0.05). It is unclear whether this effect indicates any tendency towards a confirmation bias. The reason that the teams that did not use any explicit strategy are perceived to have a lower workload than the others teams may be that not using a strategy gave them more time to focus on the actual problem.

Table 16 Average and ANOVA for the teamwork items in the observers' post-experiment questionnaire. * indicate non-overlapping ± standard error bars. ** indicate significant at p<0.05. *** indicate significant at p<0.001. ns indicate not significant.

Teamwork Item	**Critical Thinking**	**ACH**	**No Explicit**	**F(2, 12)/ F(1, 8)**	**p**
To what extent did the team members agree on how to solve the problem?	3.9	**3.5**	4.2	2.11*	0.163
To what extent did the team members present diverging views?	3.7	3.6	3.7	ns	
To what extent did the team members agree with each others suggestions?	3.4	3.4	**3.9**	5.14	0.024**
How did the dialogue of the team work?	4.0	**3.1**	4.4	4.93	0.027**
How was the team's workload?	3.7	3.4	**3.2**	2.11*	0.164
To what extent had the team sufficient	2.5	**2.2**	3.0	1.85*	0.200

time to solve the problem?					
Rank the individual team members' involvement in the experiment	4.2	**3.9**	4.4	2.71*	0.107
Rank the whole team's involvement in the experiment	4.2	**3.9**	4.4	5.43	0.021**
To what extent did the team members contribute to the solution?	7.4	8.8	**4.0**	1.31*	0.306
Describe the dominance relationship between team members.	14.2	13.7	8.8	ns	
How did the team evaluate information during the analysis?	4.1	**3.1**	4.3	9.19	0.004***
Did the team:					
a) Miss important information	20%	60%	**0%**	6.00	0.019**
b) Undervalue important information	50%	60%	**10%**	5.00	0.032**
c) Misinterpret information	20%	60%	10%	ns	
How did the team organize information in their analysis?	3.7	**3.0**	4.1	6.20	0.014**
How do you assess the team's understanding of the problem?	4.5	**3.4**	4.4	7.65	0.007***
How difficult do you think that the task was for the team?	3.3	**3.6**	3.1	3.17*	0.079
How comprehensive do you assess the team's analysis on the identified possible causes of the disease?	4.2	**3.6**	4.7	10.11	0.003***
How do you assess the team's solution?	4.1	**3.5**	4.3	10.40	0.002***
To what extent seemed members to be comfortable with the team's solution?	4.1	**3.7**	4.2	1.72*	0.220
Did the team discuss in terms of falsify or confirm hypotheses?	3.5	3.1	3.1	ns	

Table 17 Average and ANOVA for the strategy items in the observers' post-experiment questionnaire. * indicate non-overlapping ± standard error bars. ns indicate not significant.

Strategy Item	Critical Thinking	ACH	No Explicit	F(2, 12)/ F(1, 8)	p
To what extent did the team follow the analytical approach that they had received training on?	3.9	4.1	NA	ns	
To what extent did the team have sufficient understanding of the analytical strategy in order to utilize it the best?	4.0	4.3	NA	ns	
To what extent did the team focus on the following factors?					
a. Dialogue	11.5	13.5	13.5	ns	
b. Collaboration	7.5	5.5	**14.0**	2.87*	0.096
c. Valuation / assessment of information	27.5	**13.0**	25.5	1.71*	0.223
d. Structuring / organization of information	15.0	**26.0**	15.0	3.09*	0.083
e. Identification of options (reasons)	23.0	**27.0**	18.5	3.06*	0.085
f. Decision making (choice of the most plausible hypothesis)	15.5	15.0	13.0	ns	

Table 18 Percentage of the hypotheses completeness item in the observers' post-experiment questionnaire. * indicate non-overlapping ± standard error bars. ns indicate not significant.

Completeness Item	Critical Thinking	ACH	No Explicit
What causes did the team identify?			
a. Bat	100%	100%	100%
b. Zinc	100%	100%	100%
c. Medicine	100%	80%	100%
d. Flour	100%	100%	100%
e. Genetic	100%	100%	100%

3.2.5 Summary of main findings

The main findings of the experiment are that:

- Teams that use Critical Thinking perform better than teams that use Analysis of Competing Hypotheses (ACH) and the teams did not use any explicit strategy. 80% of the teams that use Critical Thinking gave the highest ranking to the most plausible hypothesis. However, the difference in performance was small compared to teams that did not use any explicit strategy.
- The critical dialogue phase of Critical Thinking was drawn out and intertwined with the information exchange for identification and clarification of hypotheses. It is therefore doubtful whether the teams actually use the strategy as intended.
- Teams that did not use any explicit strategy often mention that they would perform better using a structured strategy.
- Surprisingly, teams that use Analysis of Competing Hypotheses (ACH) performed worse than the teams that use Critical Thinking, as well as the teams that did not use any explicit strategy.
- The ACH-teams perform worse since they have difficulties using the recommended hypotheses and arguments table to get an overview of how the information supports the hypotheses. They simply had difficulty expressing the fragmented and incomplete information regarding disease agents, vectors of introduction, and disease pattern, as well as arguments and counterarguments for all the hypotheses. Instead they get trapped in confusing discussions and disagreements. Many participants perceive the ACH-strategy as complicated and more time and training may be required to apply it on complex problems.
- Although familiarity and team composition may affect the team performance, other factors are clearly more important.
- Generally, the main reason that some teams perform worse was that they had difficulty in identifying arguments and counterarguments regarding the disease pattern for whom are affected by the disease, under which circumstances, demographic differences, and the development over time.
- The teams that performed best gave the highest ranking to the most plausible hypothesis, but their arguments and counterarguments for the other hypotheses were often unsatisfactory. There is therefore considerable potential for improving the performance.

4 Discussion and conclusions

The results of the experiment show that teams that use Critical Thinking perform best, although the difference is small compared to teams that do not use any explicit analytic strategy. Further, teams that use Analysis of Competing Hypotheses (ACH) perform worst. The reason that teams using Critical Thinking perform best is that the strategy encourages team participation and openness to each other's arguments. The analysis of the team communication shows that the teams that use Critical Thinking have gained a tool to successfully focus on problem analysis and identification of plausible causes. However, there is still considerable potential for improving the performance of these teams further since the critical dialogue phase preferably should come later in the analysis. Currently, the critical dialogue is rather drawn out over a larger part of the experiment, with a communication structure that is similar to the teams that did not use any explicit strategy.

Further, the teams that did not use any explicit strategy perceive the problem as difficult and request help to structure the problem. This is not entirely consistent with the performance results and the observers' findings, which indicate that the teams have a similar performance and communication pattern as the teams that use Critical Thinking.

Finally, the teams that use Analysis of Competing Hypotheses (ACH) early on find as many hypotheses as many of the other teams, in the form of disease agents and vectors of introduction. However, they have difficulties expressing information regarding the disease pattern as arguments and counterarguments for the hypotheses. One reason for these difficulties may be that the hypotheses/arguments table, with the assumed independence between hypotheses and arguments, may not be entirely suitable for the complex problem where each possible cause represent a mental model of how different factors are combined to cause the disease. Since each possible cause then serve as a context for understanding the factors, it may be difficult to continuously change between contexts as recommended by Analysis of Competing Hypotheses to evaluate the factors against all the hypotheses before assessing their plausibility. However, the strategy is difficult to master and the performance would likely improve with more training than in the present experiment.

4.1 Concluding remarks and future work

The literature review identified both Critical Thinking and Analysis of Competing Hypotheses (ACH) as promising analytic strategies for collaborative intelligence analysis. However, the comparative evaluation of the strategies only encourages using critical thinking based on dialogue theory as an analytic strategy in collaborative intelligence analysis. Generally, the teams that use Critical Thinking seem to perform a better analysis than teams that use Analysis of Competing Hypotheses and teams that do not use any explicit strategy. Further, the results of the experiment show that teams that use Critical Thinking also have a similar communication pattern to the teams that did not use any explicit strategy and unsatisfactory quality of many hypotheses. This indicates that there is still considerable potential for improving the collaborative intelligence analysis with better analytic strategies. Preferably, such strategies should foster the participants argumentation and critical thinking skills (Fischer, Spiker, & Riedel, 2008; Kuhn, 2010), and argumentation for model development which is important for many analytic tasks (e.g. Böttcher & Meisert, 2010). Further studies are also needed of how the critical thinking strategy supports trained intelligence analysts in collaborative intelligence analysis. Such participants may well have a different set of skills due to both their professional training and higher age than the participants in the present experiment. Both education and higher age tend to enhance argumentation skills (e.g. Barchfeld & Sodian, 2009). Finally, future discourse analyses of the collected data may provide a better understanding of why the strategy Analysis of Competing Hypotheses in particular hampers how the teams use the disease pattern for their diagnostic reasoning which consequently reduces their performance.

References

Amason, A. C. & Schweiger, D. M. (1997). The effects of conflict on strategic decision making effectiveness and organizational performance. In C. K. W. De Dreu & E. van de Vliert (Eds.), *Using conflict in organizations*. London: Sage.

Barchfeld, P., & Sodian, B. (2009). Differentiating theories from evidence: The development of argument evaluation abilities in adolescence and early adulthood. *Informal Logic*, 29(4), 396-416.

Böttcher, F., & Meisert, A. (2010). Argumentation in science education: A model-based framework. *Science & Education*, Online First, 7 October 2010.

Central Intelligence Agency (2009). *A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis*. U.S. Government.

Cohen, M. S., Adelman, L., Bresnick, T. A., Marvin, F., Salas, E., & Reidel, S. (2007). Dialogue as medium (and message) for training critical thinking. In R. R. Hoffman (Ed.), *Expertise Out of Context*. Mahwah, NJ: Lawrence Erlbaum.

Costa, P. T., & McCrae, R. R. (1992). *The NEO PI-R Professional Manual*. Lutz, FL: Psychological Assessment Resources.

Elder, L., & Paul, R. (2007). *The Thinker's Guide to Analytic Thinking*. Dillon Beach, CA: The Foundation for Critical Thinking.

Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J. et al. (2005). Finding decision support requirements for effective intelligence analysis tools. In *Proc. of the Human Factors and Ergonomics Society 49th Annual Meeting*, September 26-30, Orlanda, FL. Santa Monica, CA: Human Factors and Ergonomics Society.

Fiore, S. M., Rosen, M. A., Smith-Jentsch, K. A., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors*, 52(2), 203-224.

Fischer, S.C., Spiker, V.A., & Riedel, S.L. (2008). *Training critical thinking for Army officers. Volume Two: A model of critical thinking*. ARI Research Report 1882. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Fischer, S. C. Spiker, V. A., Harris, D. H., McPeters, E. V., & Riedel, S. L. (2008). *Computerized training in critical thinking* $(CT)^2$*: A skill-based program for Army personnel*. ARI Research Report 1880. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Frey, L.R., Gouran, D.S., & Poole, M.S. (1999). *The Handbook of Group Communication Theory and Research.* Sage Publications.

Hackman, J. R., & O'Connor, M. (2005). *What makes for a great analytic team? Individual vs. Team approaches to intelligence analysis*. Washington, DC: Intelligence Science Board, Office of the Director of Central Intelligence.

Henry, R. A. (1995). Improving group judgment accuracy: Information sharing and determining the best member. *Organizational Behavior and Human Decision Processes*, 62, 190-197.

Heuer, R. J. (1999). *Psychology of Intelligence Analysis*. Washington, DC: CIA Center for the Study of Intelligence.

Heuer, R.J. (2007). Analytical tools for intelligent analysts. In *Proc. of the 8th Int. Conference on Naturalistic Decision Making*, June 4-6, Pacific Grove, CA. San Francisco, CA: San Francisco State University.

Hirokawa, R.Y., & Poole, M.S. (Eds.) (1996). *Communication and Group Decision-Making* (2nd ed.). Sage Publications.

Hutchins, S.G., Pirolli, P.L., & Card, S.k: (2006). “What makes intelligence analysis difficult?” A Cognitive Task Analysis of Intelligence Analysis. In P.L. Pirolli (Ed.), *Assisting People to Become Independent Learners in the Analysis of Intelligence* (pp. 6-52). Arlington, VA: Office of Naval Research.

Jehn, K. A. (1997). Affective and cognitive conflict in work groups: Increasing performance through value-based intragroup conflict. In C. K. W. De Dreu & E. van de Vliert (Eds.), *Using conflict in organizations*. London: Sage.

Jones, M. D. (1998). *The Thinker’s Toolkit: 14 Powerful Techniques for Problem Solving*. New York: Three Rivers Press.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*. 94(5), 810–824.

Lehner, P.E., Adelman, L., Cheikes, B.A., & Brown, M.J. (2008). Confirmation bias in complex analyses. *IEEE Transactions on System, Man and Cybernetics, Part A: Systems and Humans, 38*(3), 584-592.

Moore, D. T. (2007). *Critical Thinking and Intelligence Analysis*. Occasional Paper 14, Washington, DC: Joint Military College Intelligence Press.

Okhuysen, G.A., & Eisenhardt, K.M. (2002). Integrating knowledge in groups: How Formal Interventions Enable Flexiblity. *Organization Science, 13*(4), 370-387.

Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G.A. Klein, J. Orasanu, R. Calderwood & C.E. Zsambok (Eds.), *Decision Making in Action: Models and Methods* (pp. 327-345). Norwood, NJ: Ablex Publishing.

PARC (2009). *Sensemaking*. http://www2.parc.com/istl/groups/hdi/sensemaking/gloss-frame.htm. Palo Alto Research Centre.

Paul, R. W., & Elder, L. (2008). *The Miniature Guide to Critical Thinking Concepts and Tools* (5th Ed.). Dillon Beach, CA: The Foundation for Critical Thinking.

Pirolli, P.L., & Card, S. (2005). The sensemaking processes and leverage points for analyst technology as identified trough cognitive task analysis. In *Proc. of the 2005 Int. Conference on Intelligence Analysis*, May 2-5, McLean, VA.

Rasmussen, J., Brehmer, B., & Leplat, J. (Eds.) (1990). *Distributed Decision Making: Cognitive Models for Cooperative Work*. John Wiley & Sons.

Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes*, 88, 563-586.

Sniezek, J. A., & Paese, P. W. (1989). Improving group judgment: Increasing disagreement and reducing certainty, Paper presented at the *Annual Meeting of the Judgment and Decision Making Society*, Atlanta, GA.

Snowden, D., Klein, G., Pin, C. L., & Teh, C. A. (2007). A sensemaking experiment – Techniques to achieve cognitive precision. In *Proceedings of the 12th International Command and Control Research & Technology Symposium*. Newport, RI. 19-21 June, 2007.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: biased information sampling during discussion. *Journal of Personal and Social Psychology*, 48, 146-78.

Suthers, D. D., Vatrapu, R., Medina, R., Joseph, S., & Dwyer, N. (2007). Conceptual representations enhance knowledge construction in asynchronous collaboration. In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *The Computer Supported Collaborative Learning (CSCL) Conference 2007* (pp. 704-713). New Brunswick: International Society of the Learning Sciences.

Swedish Armed Forces (2008). *Försvarsmaktens Grundsyn Underrättelsetjänst 2008. Grundsyn Und08*. Försvarsmakten.

Straus, S. G., Parker, A. M., Bruce, J. B., & Dembosky, J. W. (2009). *The Group Matters*. Working Paper WR-580-USG. Santa Monica, RAND.

Svenmarck, P. (2011). *Dialogues as a Tool for Sensemaking in Collaborative Intelligence Analysis*. FOI Memo. Swedish Defence Research Agency.

Thorstensson, M. (2008). *Using observers for model based data collection in distributed tactical operations* (Licentiate thesis). Linköping Studies in Science and Technology 1386. Linköping, SE: Linköping University.

Trent, S., Patterson, E.S., & Woods, D.D. (2007). Challenges for cognition in intelligence analysis. *Journal of Cognitive Engineering and Decision Making*, *1*(1), 75-97.

Trnka, J., Rankin, A., Jungert, E., Lundberg, J., Granlund, R. et al. (2009). *Information Support in Modern Crisis and Disaster Response Operations*. Project report. Linköping, SE: Linköping University.

van Eemeren, F. H. & Grootendorst, R. (1992). *Argumentation, communication, and fallacies: A pragma-dialectical perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

Walton, D. N. (1998). *The new dialectic: Conversational contexts of argument*. Toronto: University of Toronto Press.

Warner, N. W, & Letsky, M. (2008). Empirical model of team collaboration: Focus on macrocognition. In M. P. Letsky, N. W. Warner, S. M. Fiore, & C. A. P. Smith (Eds.), *Macrocognition in Teams: Theories and Methodologies*. Aldershot: Ashgate.

Weick, K.E. (1995). *Sensemaking in Organizations*. Sage Publications.