# *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference*

*Human Factors in high reliability industries*

*Edited by*

***Dick de Waard, Jürgen Sauer, Stefan Röttger, Annette Kluge, Dietrich Manzey, Clemens Weikert, Antonella Toffetti, Rebecca Wiczorek, Karel Brookhuis, and Jettie Hoonhout***

**ISSN 2333-4959 (online)**

*Please refer to contributions as follows:*

[Authors] (2015), [Title]. In D. de Waard, J. Sauer, S. Röttger, A. Kluge, D. Manzey, C. Weikert, A. Toffetti, R. Wiczorek, K. Brookhuis, and H. Hoonhout (Eds.). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference (pp. pagenumbers). Downloaded from http://hfes-europe.org  (ISSN 2333-4959)

*Available as open source download*

*Published by HFES*

# Option generation in simulated conflict scenarios in approach Air Traffic Control

*Jan Kraemer[1] & Heinz-Martin Süß[2]*
*[1]DLR Braunschweig*
*[2]Otto von Guericke University Magdeburg*
*Germany*

**Abstract**

Approach air traffic controllers provide safe guidance of aircraft approaching an airport from different arrival routes. Handling traffic and preventing separation loss between aircraft requires controllers to maintain situation awareness at all times. In case an incident is foreseen, guidance options must be acquired to deal with it. Though expert controllers are expected to always come up immediately with the best guidance options, option generation skills have often been neglected in situation awareness research so far. In addition, the fact that incidents still happen in air traffic control shows the need for research in this field. In an initial investigation study, seven expert air traffic controllers completed an online-survey consisting of videos and screenshots captured from three real-time simulations of approach scenarios on Düsseldorf airport, Germany. Every scenario was designed to end in separation loss of two aircraft. In each scenario, subjects were asked to provide as many options as possible to deal with the situation one minute prior to the incident. Results showed differences between experts regarding the quality and quantity of options successfully preventing separation loss given in the scenarios, indicating different strategies of dealing with conflict situations among subjects.

**Introduction**

Air travel is considered the safest mode of transportation (IATA, 2013). However, accidents still occur and with the constant growth of air traffic over the last years, the number of incidents related to air traffic management (ATM) has also increased. Statistics revealed a number of over 120 incidents per two billion flight hours in 2012 (Eurocontrol, 2013). As recent forecasts of IATA expect a total of 3.6 billion flight passengers in 2016, about 800 million more than in 2011(IATA, 2012), the number of incidents is likely to keep growing. Highly skilled air traffic controllers are needed to manage complex traffic caused by growing numbers of aircraft and to ensure safe guidance. Safety is granted by maintaining horizontal and vertical separation between aircraft within the same sector. Additionally, compliance with limitations in altitude and flight speed must be controlled at all times. Therefore, controllers constantly have to make decisions to provide safe guidance. In 2012, there have been 125 separation minima infringements per million aircraft movements (Eurocontrol, 2013). To prevent further increase, it is important to identify the sources of human error in the decision making process.

In D. de Waard, J. Sauer, S. Röttger, A. Kluge, D. Manzey, C. Weikert, A. Toffetti, R. Wiczorek, K. Brookhuis, and H. Hoonhout (Eds.) (2015). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Decision making is a cognitive process used to find the most suitable course of action (COA) among alternatives to meet a certain goal (Wang & Ruhe, 2007). Feasible COAs are derived from careful analysis of the situation dealt with. Analysing a situation's state and figuring out what to do is called Situation Awareness (SAw; Adam, 1993). More detailed, SAw has been defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1988, p. 97). Therefore, proper SAw is necessary to select the most suitable COA from a number of possible ones depending on one's objective (Endsley, 1999a). It enables an operator to anticipate the situation's future state and to direct subsequently encoding and pattern recognition accordingly (Durso, Rawson & Girotto, 2007). Knowledge about the future state has been shown to reveal the biggest differences in SAw of experts and novices (Durso et al., 1995) and is considered to be a distinct ability of skilled experts (Endsley, 2000)

Maintaining SAw while dealing with complex dynamic situations is important for good performance. Loss of SAw has been identified as the source of operational errors in air traffic control (ATC). 58.6% of operational errors in Terminal Radar Approach Control and 69.1% in enroute ATC are caused by insufficient SAw (Endsley, 1999). As SAw involves the construction of (partially) internal representations of highly complex situations, it puts effort into cognitive resources such as working memory (Baddeley & Hitch, 1974) and attention (Durso et al., 2007). Furthermore, the dynamics of moving objects require continuous updating of those representations as the situation changes over time. SAw may be lost if complexity and dynamics exceed the capabilities of the operator's attention and working memory capacity, as both are limited resources (Endsley, 1988).

Expertise can reduce the effect of limited cognitive resources on SAw (Durso & Gronlund, 1999). Subject-matter experts develop internal models from experience which help them to guide their attention, to organise information and to project future states of the situation at hand (Endsley, 1998). Those internal models are stored in long-term memory and can be activated and integrated with situation models stored in working memory (Durso et al., 2007). As they are treated as a single piece of information, they may greatly reduce the demands of storing complex information patterns (Sweller, 2003). Sohn and Doane (2004) found that expert pilots relied on their skills built from experience when recalling flight situations from given cockpit perspectives, while novice subjects' performance was determinate by their working memory capacity.

An operator needs to know his options to adequately deal with a given situation. Confronted with familiar situations, experts are believed to come up with optimal COAs from experience without having to rely on further processing (Pfaff et al., 2013). Unfamiliar situations, on the other hand, call for more complex processing if they cannot sufficiently be mapped to prior experience. According to Wang and Ruhe (2007), setting a goal triggers an exhaustive search for possible COAs and criteria distinguishing between useful decision-strategies. This is also known as decision space (DS). It results from transforming raw information from SAw to actual COAs (Drury, Pfaff, More, & Klein, 2009). By analysing the potential costs and outcomes of available options, operators eventually decide on the most suitable one. In emergency response decision making research, assistance systems have been developed to help decision makers to find available COAs and compare them in terms of possible outcomes (Chandrasekaran, 2007). Using exploratory algorithms, such systems are designed to reduce the effects of uncertainty and time pressure in complex dynamic situations on the operator's performance. They provide the decision maker with all possible COAs, their respective outcomes and possible risks and robustness over a variety of conditions (Pfaff et al., 2013). It has been shown that such systems can improve both the accuracy and speed of identifying robust decisions from a set of alternatives (Lempert, Popper & Banks, 2003, cited by Pfaff et al., 2013).

Constructing the DS of an operator requires proper SAw and involves knowledge and expertise. SAw is affected by limited cognitive resources and must be maintained at all times to handle complex dynamic situations. Furthermore, decisions must often be made under time pressure. Given unlimited time to analyse a situation without having to memorise all the details, subject-matter experts should be able to build up sufficient SAw to deal with the situation. Thus, in combination with their expertise, they should be able to provide an enclosing set of possible COAs. In highly standardised and regulated fields such as ATC, DS are expected to bear a close resemblance among experts, because explicit rules can put an external limit to the possible options a decision maker has to find and compare. The aim of this study was to find out if experts are actually able to generate encompassing DS if they have both unlimited time and access to all relevant information. Under these conditions experts are believed not to differ in conflict resolution performance among each other. Thus, no significant differences between experts are expected in terms of both quality and quantity of their decisions.

## Methods

### *Subjects*

Ten approach and one tower air traffic controller (10 male, 1 female) from Deutsche Flugsicherung (DFS) participated in the experiment. Age ranged from 23 to 51 years ($M$ = 32.82, $SD$ = 8.55) while years of experience ranged from 2 to 20 years ($M$ = 8.36, $SD$ = 6.23). Subjects were recruited by direct advertisement via the internal network of DFS. Participation was voluntary, no expense allowance was paid.

### *Conflict resolution task*

Subjects were asked to provide as many solutions as possible for conflict scenarios in simulated approach ATC. A computer based survey was created containing three short real-time simulated scenarios of approaching air traffic on Düsseldorf airport (EDDL), Germany. Scenarios were created using NLR ATM Research Simulator (NARSIM; ten Have, 1993), a real-time ATM simulator software developed by the National Airspace Laboratory of the Netherlands. Scenarios showed aircraft approaching Düsseldorf Airport via Standard Arrival Routes using conventional Transition To Final procedures (see Figure 1). Scenarios each lasted between four and five minutes and were designed to end in separation loss between two aircraft.

Videos of the simulation runs were recorded using desktop capturing software. Additionally, screenshots of the situation one minute prior to separation loss were captured with the respective aircraft highlighted.

Each item was introduced by a short description of the situation at hand including the time at which the conflict occurred as well as the conflicting aircraft. Underneath the introduction, the video and the screenshot of the current conflict scenario were embedded. Subjects were asked to watch the videos as well as the screenshots carefully and to find as many solutions as possible to prevent the upcoming separation loss one minute before it occurred. Separated pre-labelled tables were presented on the same page to write down advisories that would be given to the aircraft. All advisories written in one table represented one COA. Subjects were allowed to advise changes to flight speed and altitude of aircraft and to turn aircraft from the downwind to the centreline. Additionally, subjects were asked to rate if they would personally use each option in reality on a Likert scale ranging from 1 (never) to 7 (absolutely). Subjects were allowed to watch the videos and screenshots as often as they wanted and to switch back and forth between the scenarios to find as many options as possible.

Several simplifications were set in the simulation to make answers more comparable. All aircraft were set to the same type. No differences in horizontal separation had to be considered between different wake turbulence categories and no wind was present. All aircraft followed the approach procedure as described.

*Procedure*

The conflict resolution task was presented as an online survey. First, a biographical questionnaire was completed. Following the questionnaire, general instructions were presented, involving information about aircraft types, conventional approach procedure, how to change video settings and to fill out the direction tables. Furthermore, subjects were assured that no data could later be linked to a specific person. Thereafter subjects completed the questionnaires as described.

*Data analysis*

Validation of the options provided by the participants was done by replaying the scenarios once for each answer. One minute prior to the separation loss, the simulation was paused and all advisories for one solution were put into the simulation. If the separation loss was prevented successfully, the respective option was scored with one point. If any violations of limitations to speed and altitude were made, half a point was given. Zero points were given if the conflict still occurred or new conflicts were produced. Options were categorized by combinations of directions given. This way, small deviations in absolute values assigned between subjects did not count as distinct options.

## Results

Subjects provided a total of 12 original options in total for scenario one. Ten options were found in scenario two and seven in scenario three. Descriptive statistics of valid options provided per subject in each scenario are presented in Table 1. No significant deviations from either uniform or normal distribution were found using Kolmogorov-Smirnov-Tests in any scenario. Paired t-tests showed significant deviations of mean numbers of valid options per subject from total valid options provided by all subjects in scenario one ($t(10) = -28.03$, $p < .001$), two ($t(10) = -33.80$, $p < .001$) and three ($t(9) = -21.10$, $p < .001$).

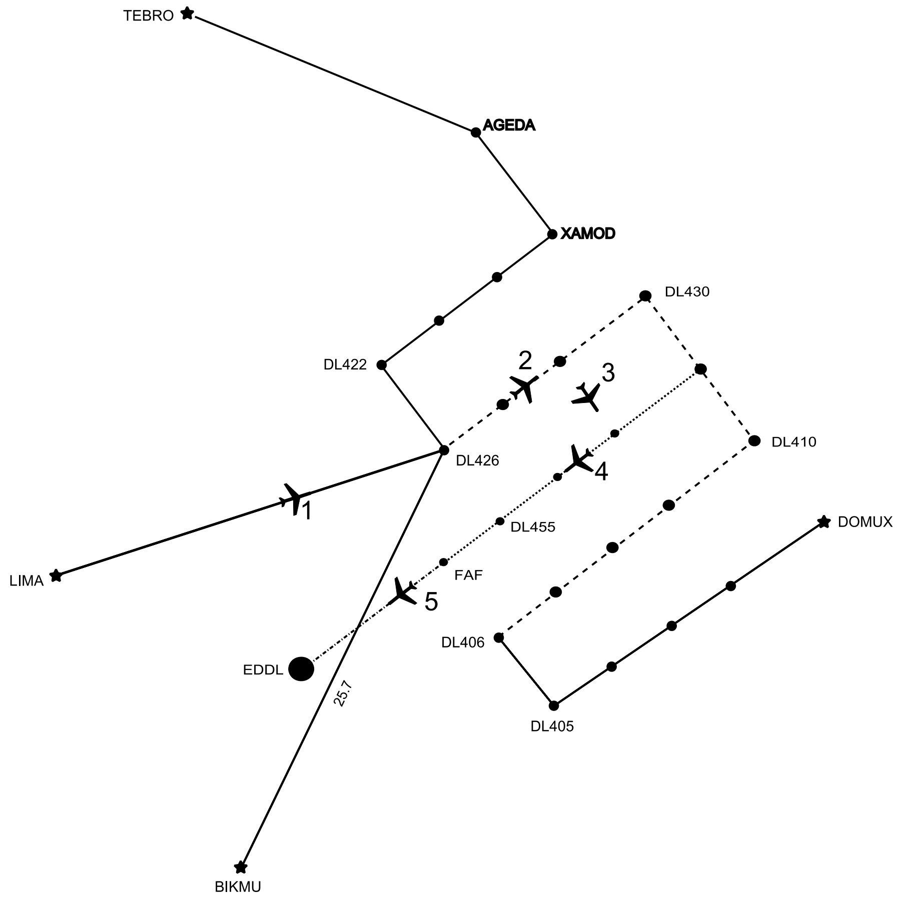


*Figure 1. Schematic depiction of conventional Transition To Final procedure at Düsseldorf Airport (EDDL), Germany*

In order to take differences in experience into account, subjects were divided into two groups by median split (*Mdn* = 6). A 3x2 ANOVA with scenarios as within-subject factor and experience (low vs. high) as between-subject factor revealed significant differences in the total number of options provided per subject among scenarios ($F(2, 18) = 7.43$, $p = .005$, $\eta^2 = .48$). Post-hoc Bonferroni-corrected paired t-tests showed no significant differences between scenarios. No significant effects of scenarios ($F(2, 18) = 2.87$, $p = .086$) or experience ($F < 1$) on mean numbers of valid options were found. No significant correlation between experience and valid options were found throughout scenarios ($r_s$ = .22, $p$ = .257*).*

*Table 1. Distribution of valid options provided per scenario (Top – Total number of valid options provided over all subjects, % valid – percentage of valid answers given, KS-Z Kolmogorov-Smirnov-Z values of tests for uniform distribution)*

| Scenario | Top | *MD* | *SD* | % valid | Min. | Max. | *KS-Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | 12 | 2.0 | 1.2 | 70.9 | 0.0 | 4.0 | 0.53 | .943 |
| 2 | 10 | 1.2 | 0.8 | 46.2 | 0.0 | 2.5 | 0.63 | .819 |
| 3 | 7 | 1.5 | 0.8 | 72.5 | 0.0 | 3.0 | 0.74 | .648 |

Frequencies of common options provided by subjects were counted for each scenario (see Figure 1). Out of the 12 options in scenario one, four options were stated by more than two participants. One option out of ten in scenario two and three out of seven options in scenario three were used more than twice.

No significant correlation between ratings and validity of options were found among scenarios ($r(75)$ = -.05, $p$ = .348). Paired t-tests between mean ratings of valid and invalid options showed significantly higher ratings of valid options in scenario two ($t(5)$ = 4.72, $p$ = .005, $d$ = 1.42). No significant differences of mean ratings were found in scenario one ($t(2)$ = -0.28, $p$ = .808, $d$ = -0.17) and three ($t(3)$ = -2.85, $p$ = .065, $d$ = -1.88) Out of the eight common options used by three or more subjects, ratings of four solutions differed no more than two points. Ranges in the remaining options went up to a maximum of five points.

**Discussion**

It was expected that various subject-matter experts would provide highly similar sets of possible COAs when confronted with the same conflict scenarios. Data showed that this is not the case, even though Kolmogorov-Smirnov-Tests showed no significant differences in quantity of given and valid answers. As no deviation from standard distribution was found as well, the results indicate that the tests lacked significance due to the low power arising from the small sample. Various experts came up with a lot of different solutions to the same situation. Furthermore, none were even close to providing all possible solutions in any scenario. While a total of 12, ten and seven different options were given in total, the maximum number of experts sharing one option was never higher than four among scenarios. Moreover, some high differences between ratings for the same options were found among subjects. This is surprising considering that all of the participants were highly trained professional air traffic controllers. Insufficient SAw as a cause of error was unlikely as unlimited time was given to solve each scenario and all relevant information was accessible throughout the task. Additionally, no information had to be memorized over longer periods of time because videos and screenshots could be watched repeatedly. Nevertheless, subjects not only failed to provide complete DS but even produced invalid solutions which did not solve the respective conflicts.

As subjects' experience covered a range of 18 years, this might explain differences in DS. As experience and knowledge were discussed as important factors in acquiring SAw, an increasing number of correct answers would be expected with higher experience. However, the data do not support this explanation as no higher scores were found for the more experienced subjects. Nevertheless, it should be kept in mind that this might stem from the small sample, namely the lack of statistical power as mentioned. In a larger sample, experience might make a difference when it comes to finding solutions in emergency situations.

The available advisories and simplifications used during the experiment may pinpoint another explanation for the differences in experts' DS. Some subjects criticised the lack of heading advisories claiming that this eliminated possible options. In that case, experts should have been even more likely to produce similar sets of COAs due to the reduced amount of options left. The low level of compliance found among the answers provided throughout the experiment contradicts this point. Although options were excluded from the start, subjects still came up with a lot of different approaches to the same problems and differed strongly in both quantity and quality of their answers. Allowing for more directions might have resulted in even bigger variance of both. Unfortunately, it was not possible to test this supposition with the acquired data.

The low number of options may result from subjects tending to provide only robust COAs instead of encompassing DS. It has been argued that optimal COAs are almost impossible to find in complex dynamic situations due to their high levels of uncertainty and time pressure (Lempert et al., 2003, cited by Pfaff et al., 2013). Therefore, decision makers tend to make robust decisions which maintain their effectiveness over a wider range of possible outcomes and conditions in emergency situations (Bryant & Lempert, 2010). However, two findings in this experiment contradict this explanation. First, subjects were only watching a simulation and were given as much time as they wanted to produce their answers. Therefore, it is unlikely that time pressure kept them from thinking all their options through or rushed them towards making decisions. Second, subjects also provided options rated with only two points (very unlikely), meaning they gave an answer they would not really use in a real-world situation.

Subjects may have provided fewer answers than they could possibly have due to lack of motivation. As the task required them to rethink a situation over and over to come up with new ideas, this might have reduced compliance with the task over time even though participation was voluntary. Indeed, the descending number of total options provided per subject among scenarios indicates loss of motivation throughout the task. On the contrary, no decrease in valid options was found between scenarios. Loss of motivation may explain why fewer answers were provided in the last scenario. However, it does not explain why the quality of the answers did not drop over scenarios. Due to anonymity, contacting participants in order to confront them with the results and ask about problems afterwards was impossible. Future studies of this kind could be combined with post experimental interviews to allow for more detailed explanations of strategies used to identify possible COAs. Additionally, allowing subjects to further explain their answers might help to keep up motivation throughout the task.

It has been argued that current air traffic systems will not be able to cope with projected increases in air traffic due to lack of flexibility (Lohr & Williams, 2008, cited by Pfaff et al., 2013). Assistance systems which have been developed and are already in use by some air navigation service providers may help air traffic controllers to overcome this problem by providing a broader range of COAs (Pfaff et al., 2013). Looking at the data, the question arises if such systems should already be mandatory for emergency decision making in ATC. Although unlimited time was given to solve each scenario, subjects still produced invalid answers which didn't prevent the conflicts. In addition, in each scenario at least one subject failed to produce any valid options at all. In future studies, it would be interesting to compare the DS of human experts directly to emergency assistance systems which make use of robust decision making processes. If all possible COAs and their estimated outputs were derived from modelling processes, it could be tested if human experts are able to provide a similar set of answers. Additionally, it could be examined if DS of human experts, although they may be smaller in quantity, are representing the most robust COAs found by the assistance systems. Unfortunately, such systems were not available in this study. Furthermore, although it has been argued that lack of SAw was an unlikely cause of error in this study, this cannot be ruled out. Future work should include the assessment of SAw data using probe methods such as the Situation Present Assessment Method (Durso, Blackley & Dattel, 2006) to draw more resilient conclusions about Saw and DS generation.

## References

Baddeley, A.D., & Hitch, G.J. (1974): Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Bryant, B., & Lempert, R.J. (2010). Thinking Inside the Box: A Participatory, Computer Assisted Approach to Scenario Discovery. *Technology Forecasting and Social Change, 77*(1), 34-49.

Chandrasekaran, B. (2007). *From Optimal to Robust COAs: Challenges in Providing Integrated Decision Support for Simulation-Based COA Planning*. Laboratory for AI Research, The Ohio State University.

Drury, J.L., Klein, G.L., Pfaff, M., & More, L. (2009). Dynamic decision support for emergency responders. In *Proceedings of the 2009 IEEE Technologies for Homeland Security Conference* (pp. 537 – 544). Waltham, USA: Technologies for Homeland Security.

Durso, F.D., Blackley, M.K., & Dattel, A.R. (2006). Does Situation Awareness Add to the Validity of Cognitive Tests? *Human Factors, 48*, 721 – 733.

Durso, F.D., & Gronlund, S.D. (1999). Situation Awareness. In F.T.R.S. Nickerson, S.T. Dumais, S. Lewandowsky, and T.J. Perfect. (Eds.), *Handbook of Applied Cognition* (2nd ed., pp. 163-193). Chichester, UK: Wiley.

Durso, F.T., Truitt, T.R., Hackworth, C., Crutchfield, J., Nikolic, D., Moertl, P (1995). Expertise and chess: A pilot study comparing situation awareness methodologies. In D.J. Garland & M.R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 295–304). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Endsley, M. (1988). Design and Evaluation for Situation Awareness Enhancement. In *Proceedings of the Human Factors and Ergonomics Society 32nd Annual Meeting* (Vol. 1, pp. 97 – 101). Anaheim, CA: Human Factors Society.

Endsley, M. (1999). Situation Awareness and Human Error: Designing to Support Human Performance. In *Proceedings of the High Consequence Systems Surety Conference*. Albuquerque, NM: Sandia National Laboratory.

Endsley, M. (1999a). Situation Awareness In Aviation Systems. In D.J. Garland, J.A. Wise, and V.D. Hopkin, (Eds.), *Handbook of Aviation Human Factors* (pp. 257-276). Mahwah, NJ: Lawrence Erlbaum Associates.

Eurocontrol (2013). *Annual Safety Report 2013.* Retrieved from http://www.eurocontrol.int/sites/default/files/content/documents/single-sky/src/src-docs/src-doc-53-e1.0.pdf

Gheisari, M., & Irizarry, J. (2011). Investigating Facility Managers' Decision Making Process through a Situation Awareness Approach. *International Journal of Facility Management, 2*, 1-11.

International Air Transport Association (2013). *IATA Annual Review 2013*. Retrieved from http://www.iata.org/about/Documents/iata-annual-review-2013-en.pdf

Pfaff, M.S., Klein, G.L., Drury, J.L., Moon, S.P., Liu, Y., & Entezari, S. (2013). Supporting Complex Decision Making Through Option Awareness. *Journal of Cognitive Engineering and Decision Making, 7,* 123-140.

Sohn, Y.W., & Doane, S.M. (2004). Memory Processes of Flight Situation Awareness: Interactive Roles of Working Memory Capacity, Long-Term Working Memory, and Expertise. *Human Factors, 46,* 461-475.

Sweller, J. (2003). Evolution of Human Cognitive Architecture. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 43, pp. 215-266). San Diego, CA: Academic Press.

Wang, Y., & Ruhe, G. (2007). The Cognitive Process of Decision Making. *International Journal of Cognitive Informatics and Natural Intelligence, 1*, 73-85.