Determinants of Conflict Detection: A Model of Risk Judgments in Air Traffic Control

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Human Factors: The Journal of Human Factors and E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>06-2328.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Regular Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>decision making, naturalistic decision making &amp; COGNITIVE PROCESSES, INDIVIDUAL DIFFERENCES, anticipation, expertise</td>
</tr>
</tbody>
</table>
Determinants of Conflict Detection:
A Model of Risk Judgments in Air Traffic Control

Stéphanie Stankovic and Éric Raufaste
CNRS and University of Toulouse, Toulouse, France

Philippe Averty
DGAC/DSNA/DTI/SDER, Toulouse, France

Keywords: individual differences, anticipation, expertise
Abstract

**Objective:** A model of conflict judgments in air traffic control (ATC) is proposed. **Background:** Three horizontal distances determine risk judgments about conflict between two aircraft: (i) $D_{t0}$ is the distance between the crossing of the aircraft trajectories and the first aircraft to reach that point; (ii) $D_{th}$ is the distance between the two aircraft when they are horizontally closest; (iii) $D_{tv}$ is the horizontal distance between the two aircraft when their growing vertical distance reaches 1000 feet. **Methods:** Two experiments tested whether the variables in the model reflect what controllers do. In Experiment 1, 125 certified controllers provided risk judgments about situations where the model variables were manipulated. Experiment 2 investigated the relationship between the model and expertise by comparing a population of certified controllers with a population of ATC students. **Results:** Across both experiments, the model accounted for 44-50% of the variance in risk judgments by certified controllers ($N = 161$), but only 20% in judgments by ATC students ($N = 88$). There were major individual differences in the predictive power of the model as well as in the contributions of the three variables. In experiment 2, the model described experts better than novices. **Conclusion:** The model provided a satisfying account of the data, albeit with substantial individual differences. It is argued that an individual-difference approach is required when investigating the strategies involved in conflict judgment in ATC. **Application:** These findings should have implications for developing user-friendly interfaces with conflict detection devices and for devising ATC training programs.
Introduction

Understanding how experts become aware of the potential risks associated with a given situation is a major issue in cognitive ergonomics. In this respect, air traffic control (ATC) is a particularly interesting situation, as experimental stimuli can be carefully controlled and even laypeople can understand the potential costs of an error. In the past, researchers have concentrated on modeling the ATC task as a whole (e.g., Bisseret, 1981; Davison et al., 2003; Histon et al., 2002; Niessen et al., 1999; Seamster et al., 1993; Wickens et al., 1997) but there has recently been an upsurge of interest in a more reductionist approach, focusing solely on conflict detection (Averty, 2005; Bisseret, 1995; Rantanen & Nunes, 2005; Xu & Rantanen, 2003).

The air traffic control task

The ATC task basically consists in making decisions in order to allow fluid traffic, all the while avoiding potential risks of conflict (Abdesslem et al., 1999). Thus, the crux of the air traffic controller’s job in approach control is sequencing the aircraft with an orderly and regular spacing, and maintaining a safe separation between aircraft (Wickens et al., 1997).

The concept of “conflict” between aircraft

Technically, an “air proximity” or "airprox" occurs when the distance between two aircraft falls below thresholds conventionally fixed at 3 or 5 nautical miles (nm) on the horizontal plane, and 1000 feet on the vertical plane. Conflict detection is the process whereby controllers identify potential airproxes, by anticipating aircraft future positions. A potential airprox cannot always be detected as soon as an aircraft arrives on the radar screen, as the situation may change unpredictably, but detection must start as soon as an aircraft is...
announced, even before it appears on the screen (e.g. there may be a potential conflict between a take-off and a low-altitude flight).

The data processed by ATCs

The present study focused on approach control, in which controllers have to pay particular attention to changes in altitude. To do so, they mainly rely on radar information. ATCs have to perform a complex task involving huge amounts of data, of a highly diverse nature (concerning not only traffic but also their working environment). These parameters are dynamic (their values change over time, even without any human intervention), fuzzy (there is often considerable imprecision in the data) and uncertain (e.g. some devices give an estimate of an aircraft’s future position at, say, time $t$, but it is impossible to know with certainty what the aircraft’s actual position will be at $t$). They may even be missing or incomplete (not all flight parameters and pilots’ intentions are available to the controller).

Cognitive processes in ATC

Relevance of the mental representation

Due to the complexity of the ATC task and the effects of time pressure on decision-makers (Fennema & Kleinmuntz, 1995; Payne et al., 1993), a central question concerns the quantity and type of data that will actually be taken into account. For example, Payne reported that as time pressure increases, decision-makers focus their attention on the most important information (Payne et al., 1988). Moreover, reducing task complexity is achieved by reducing the total amount of information that has to be taken into account, with priority being given to the most relevant items (Niessen et al., 1999). All controllers build mental representations, but their particular conception of relevance may influence the way they take in and memorize information (Durso et al., 1998a). Controllers select the most relevant items from the data they have to hand,
Risk judgments in ATC

such as altitude and route (Boudes, 1997; Gronlund et al., 1998; Willems et al., 1999). Their “flight situation awareness” (Endsley, 1988) relies on the retention of elements ranked according to their importance, and Gronlund et al. (1998) found more flight data about “important” aircraft in working memory than data about aircraft deemed to be less important.

Overview of conflict detection processes

Lafon (1978) found that the majority of controllers estimate separation between aircraft on the horizontal plane, and while others use the vertical plane, few use both. Other descriptions of air traffic controllers’ cognitive activity have been produced in the last 10 years (Boudes & Cellier, 1998; Histon et al., 2002; Niessen et al., 1999; Wickens et al., 1997), in terms of data selection, anticipation, conflict resolution and control.

A recent study of en route control at constant altitude suggested that ATCs use a lexicographic method for appraising conflict risks (Rantanen & Nunes, 2005). They begin by comparing aircraft altitudes. If there is no potential conflict, the process ends there. If the difference in altitudes falls below the conventional 1000 feet threshold, horizontal distances and trajectories are used to estimate the minimum horizontal distance. A conflict is detected when the minimum horizontal distance falls below the conventional threshold of 3 or 5 nautical miles (nm). This strategy can scarcely be applied, however, in a situation where vertical separation is so crucial. Averty (2005) emphasized the importance of processing the spatiotemporal information available on the interface, and the model is presented here in a revised version:

- A first variable, referred to here as $D_{t_0}$, corresponds to situation at the present time ($t_0$). $D_{t_0}$ is the horizontal distance between the crossing of the trajectories and the first aircraft to reach that point (Figure 1, top).
The second variable, known here as $D_{th}$, corresponds to the anticipated time ($t_h$) when the aircraft will be closest on the horizontal plane. $D_{th}$ is the minimal horizontal distance (ground projection). This variable was also used in Rantanen and Nunes’ model (Figure 1, center) (2005).

Let us call $t_v$ the time when the two aircraft reach a growing vertical separation of 1000 feet. The third variable, $D_{tv}$, corresponds to the horizontal distance (ground projection) at $t_v$ (Figure 1, bottom).

Depending on the situation, $t_h$ may come before or after $t_v$. The rationale for choosing these variables is that when taken together, the parameters $D_{th}$ and $D_{tv}$ correspond to the exact definition of conflict. $D_{th}$ is similar to CPA (“closest point of approach”), a well-known variable in the ATC literature (Rantanen & Nunes, 2005; Xu et al., 2004), while $D_{t0}$ is based on the concept of “relative judgment” (RJ), which has also been defined in previous studies (Law et al., 1993; Treselian, 1995). RJ consists in estimating which aircraft will be the first to reach the crossing point. According to Xu and Rantanen (2003), RJ is a subtask of conflict detection.

**Cognitive analysis of the model**

From a cognitive standpoint, the variables entail different cognitive resource levels. $D_{t0}$ is the distance between the crossing point and the aircraft that will reach the crossing point first. Thus, taking $D_{t0}$ into account might involve mentally inferring the intersection of the trajectories, which provides one distance for each aircraft. $D_{t0}$ is therefore the distance associated with the aircraft with the highest ratio of speed vector to distance from crossing point. $D_{t0}$ is usually easy to assess because finding the highest ratio can be straightforward in many situations (e.g. both aircraft have one parameter in common and one that is different: approximately the same speed but one is clearly closer, or about same distance from the intersection but one is clearly faster).
Estimating $D_{th}$ is harder because it requires mentally moving the two aircraft along the horizontal plane according to their respective speeds, and inferring the time when they will come closest on that plane. Here, the aircraft’s positions at that point have to be maintained mentally in order to estimate the distance between them. $D_{t}$ is the most complex variable to compute, as the two aircraft first have to be mentally moved according to their respective vertical speeds (altitude information alone is provided, in numerical form). The controller then has to establish the point at which the vertical separation is growing and reaches the 1000 feet threshold. Lastly, the horizontal positions of the two aircraft have to be inferred and maintained at that point, in order to estimate the distance between their positions. Anticipation is known to be a difficult mental activity, with experts having a clear and well-documented advantage (e.g., Boudes & Cellier, 1998; Denecker, 1999). We therefore predicted that the model’s variables would contribute to risk judgments.

Experiment 1

Hypotheses

This study was actually a re-analysis of data initially collected by Averty (2005). The purpose of the present study was to estimate the ability of the model’s variables to predict risk judgments by ATCs. According to our first hypothesis (H1), the model as a whole would account for a substantial proportion of the experienced controllers’ judgments about the risk of conflict. Moreover, each of the three variables ($D_{t0}$, $D_{th}$ and $D_{t}$) would significantly contribute to the prediction of risk judgments. Given the definitions of $D_{th}$ and $D_{t}$, we expected these variables to contribute negatively to risk judgments (H2a and H2b) because the probability of a conflict diminishes as these distances increase. Predictions relating to $D_{t0}$ are more complex because this variable is not sufficient for making a decision. Whether or not it affects risk judgments depends
on how this distance is interpreted. Two contrasting patterns are possible. According to the first interpretation, higher values of $Dt_0$ correspond to a situation where determining a conflict becomes harder as the distances between the crossing point and the aircraft increase. Hence, higher values of $Dt_0$ may increase the feeling of insecurity and risk judgments. In such cases, we would expect to find a positive relationship between $Dt_0$ and risk judgments (H2c). Another possible interpretation is that situations where $Dt_0$ has high values simply offer more time before a clear decision needs to be made. Such an interpretation would lead to a negative contribution (H2d).

Lastly, the conjunction of vertical and horizontal separation is not always necessary to decide for or against a conflict. If there is sufficient separation along one of the planes, this will be enough to decide that there is no conflict. Hence, it is possible that some controllers get into the habit of focusing on one dimension first and other controllers on the other dimension. Moreover, some controllers may prefer to look for conflict first, whereas others may prefer to look for non-conflict first. For exploratory purposes, we also investigated individual differences in weighting patterns, as these results might offer insights into individual strategies.

Method

Participants. A total of 125 controllers (73.6% men, 26.4% women) from three different airports (Lyons, $N = 45$; Marseilles, $N = 45$; and Toulouse, $N = 35$) volunteered to take part. Their ages ranged from 26 to 56 ($M = 43.17$ yrs, $SD = 8.72$ yrs). The length of experience after certification in their current sector was known for 114 of the 125 participants and ranged from 1 to 25 years ($M = 8.64$ yrs, $SD = 6.73$ yrs).

Variables. The three variables in the model were manipulated by means of scenarios featuring real radar recordings of aircraft trajectories. $Dt_h$ was given values that were chosen
after examining a database of 150 conflicts (Averty, 1998), which showed that solutions to these conflicts created a 12-nm separation on average (context of radar approach traffic). To keep the sample size at a realistic level, five values were selected: 0, 3, 6, 9, or 12 nm. \( D_t \) could take values ranging from 0 to approximately 41 nm, i.e. ten minutes of flight at the usual initial approach or take-off speed. Given ATC rules and sector sizes, detecting conflict is irrelevant beyond 41 nm. We then gradually segmented this maximum interval in order to define the different levels of \( D_t \). \( D_v \) could take the values 0, 5, 10, 15, or 20 nm. These levels were chosen by taking into account the 2D projection at the point of minimum separation (1000 feet). Each participant ran 75 scenarios. The main DV was the risk judgment itself, provided on a Likert 8-point scale ranging from certainty of a conflict furthest left to “no problem” furthest right. Each button was assigned a value, so as to produce 8 values ranging from -3.5 to 3.5. Positive values (> 0) reflected greater feelings of risk, whereas negative values (< 0) reflected greater feelings of safety.

**Material and procedure.** The experiment lasted about 30 minutes. Each participant underwent 4 familiarization trials, followed by 75 experimental trials in random order. In each one, the radar screen showed a pair of converging aircraft in a given sector, one at a stable altitude, the other either climbing or descending. The aircraft followed trajectories that are actually used around the airport where the data were collected and participants could see the strips giving the aircraft’s type and flight plan. Speed vectors were displayed with a 3-min. time span. Participants had to estimate the risk of conflict. The radar screen, updated every 4 seconds, displayed the horizontal speed, flight reference and vertical speed.

**Analysis.** Three intra-individual betas (\( \beta-D_t \), \( \beta-D_t \) and \( \beta-D_v \)) were obtained, using multiple linear regressions across the 75 scenarios. Each beta represented the contribution of the
predictor \((D_{t_0}, D_{h}, \text{and} \, D_{tv})\) to the risk judgment. The overall fit of the model for a participant was given by the adjusted \(R^2\). Absolute contributions were tested using one-sample \(t\) tests on betas, with 0 as the reference value. Relative contributions were tested using multivariate ANOVAs on betas. To report effect sizes, we used the adjusted \(R^2\) for ANOVAs, and Cohen’s \(d\) for mean comparisons, with Cohen’s conventions as a reference (Cohen, 1988): the effect was deemed to be small if \(R^2 \in [0.01, 0.09]\) or \(d \in [0.2, 0.5]\), medium if \(R^2 \in [0.09, 0.25]\) or \(d \in [0.5, 0.8]\) and large if \(R^2 \geq 0.25\) or \(d \geq 0.8\). Lastly, a cluster analysis was performed in order to identify potential strategy differences. The classification was made using the K-means method in Statistica (StatSoft, 2005), asking for 2 clusters. A multivariate ANOVA was then used to compare the clusters. Comparisons between the mean \(R^2\) of the two clusters were made using the Mann-Whitney U test. Correlations between the betas within each cluster and the length of experience were also processed. Throughout the paper, the statistical significance threshold was .05.

**Results and Discussion**

**Overall fit of the model.** Descriptive statistics are given in Table 1. Intra-individual regressions confirmed that the model as a whole captured a non-negligible proportion of the variance in risk judgments about conflict (more than 44% of explained variance), thus validating our first hypothesis. However, there was considerable intraindividual variability in the proportions of variance explained by the model: some controllers did not correspond to the model at all, whereas for others the model explained 80% of the variance (Figure 2).

*Insert Table 1 about here*

All the variables in the model made significant contributions. As expected in hypothesis H2a, \(D_{h}\) made a negative contribution \((t(124) = 18.08; \, p < .001)\), with a large effect \((d = 1.62)\). As expected in hypothesis H2b, \(D_{tv}\) also made a negative contribution \((t(124) = 26.53; \, p < .001)\),
Risk judgments in ATC

with a large effect ($d = 2.37$). Thus, the higher the values of $D_{th}$ or $D_{tv}$, the lower the feelings of risk. For all participants, $D_{t0}$ contributed positively to risk judgments ($t(124) = 7.81; p < .001$), with a medium effect ($d = 0.70$). This is compatible with the interpretation where $D_{t0}$ conveys uncertainty (H2c), but not with the interpretation where $D_{t0}$ conveys a feeling of available time (H2d). In other words, high values of $D_{t0}$ may have triggered a feeling of insecurity because with higher values of $D_{t0}$ apparent distances are greater, making it more difficult to estimate the potential for conflict.

Interestingly, the weight of $D_{tv}$ was negatively correlated with the two others: between $\beta - D_{tv}$ and $\beta - D_{th}$, $r(124) = -.64, p < .001$ and between $\beta - D_{tv}$ and $\beta - D_{t0}$, $r(124) = -.23, p = .010$. There was also a significant positive correlation between $\beta - D_{tv}$ and $\beta - D_{t0}$, $r(124) = .42, p < .001$. Thus, the more $D_{tv}$ reduced risk judgments, the more the other two variables increased risk judgments. We propose the following interpretation: because only one variable of $D_{tv}$ and $D_{th}$ can be decisive, the more $D_{tv}$ is used, the less $D_{th}$’s contribution is needed. This competition between the variables may explain the strong negative correlation observed between $\beta - D_{tv}$ and $\beta - D_{t0}$. No clear explanation appears for correlations involving $D_{t0}$.

**Individual differences in risk judgments about conflicts.** Two patterns of contributions were identified. We chose to refer to them as the “majority pattern” ($N = 76$) and the “minority pattern” ($N = 49$).

**Insert Figure 2 about here**

On average, the model explained a significantly higher proportion of variance in participants displaying the majority pattern than in participants displaying the minority one (46% vs. 41%, $z = 2.05, p = .040$). Within each pattern, the three betas were significantly different from 0. In the majority pattern, there were medium-to-large effects, $t_s(75) > 3.51, p_s < .001, d_s$.
Risk judgments in ATC

> 0.4, whereas in the minority pattern, all effects were large, $t_s(48) > 6.61; p_s < .001; d_s > 0.94$. Both patterns exhibited positive contributions of $Dt_0$, and negative contributions of $Dt_h$ and $Dt_v$.

Thus, one important result is that neither pattern contradicted our H1a, H2a, H2b and H2c hypotheses.

Insert Figure 3 about here

In order to interpret the differences in the patterns, a multivariate ANOVA was performed, with the patterns as IVs and the three weights as DVs. Post-hoc tests showed (1) more negative $\beta$-$Dt_0$s in the minority pattern than in the majority one ($p < .001$); (2) more positive $\beta$-$Dt_0$s and less negative $\beta$-$Dt_h$s in minority-pattern participants than in majority-pattern ones ($p_s < .001$). This confirms that the two clusters corresponded to two different populations in terms of strategies (Figure 3).

One could still argue that these two patterns only captured noise, and it is difficult to put forward conclusive arguments against that interpretation. However, several observations deserve attention:

1. A significant negative correlation was observed between $\beta$-$Dt_0$ and $\beta$-$Dt_h$ in the majority pattern ($r(75)=-.43, p < .001$), but not in the minority pattern ($r(48) = -.16, ns$). This is compatible with an interpretation where majority participants decide more on the basis of the horizontal separation and therefore do not devote much energy to computing the vertical separation. In the minority pattern, none of the correlations between the three betas remained significant, though this may have been due to a lack of statistical power, as the correlation between $Dt_0$ and $Dt_h$ still reached $r(48) = .26$.

2. Participants in the two patterns did not differ significantly with regard to their experience in the sector, $M = 9.03$ yrs ($SE = 0.90$ yrs) vs. $M = 8.09$ ($SE = 0.83$ yrs), $t(112) = 0.74$, $ns$. 
ns. However, in the minority pattern, there was a positive correlation between experience in the sector and $\beta-Dt_h$ ($r(46) = .40, p = .005$), but no correlation at all with the two other betas ($r(46) < .11, ns$), whereas in the majority pattern, there was no correlation between experience and $\beta-Dt_h$ ($r(66) = -.01, ns$), but significant correlations with $\beta-Dt_v$ ($r(66) = .33, p = .006$) and $\beta-Dt_0$ ($r(66) = -.26, p = .030$). One possible interpretation is that (1) some participants were captured better by $\beta-Dt_v$ and others by $\beta-Dt_h$, and (2) less experienced participants were described better by the model than more experienced ones.

**Conclusion of Experiment 1**

Overall, the model produced a satisfactory account of the data. Some of the results are troubling, as we would have expected the model to fit the data from the more experienced controllers better than the data from the less experienced ones, instead of which, the correlations within each of the two clusters revealed a pattern where the model’s predictions tended to be countered by experience in the sector. This result may simply be an artifact stemming from a general attitude of cautiousness in the more experienced controllers – an attitude that leads them to complement the strategies captured by the model by other kinds of strategies. If this is the case, we would expect to observe a steady improvement in the model’s predictive power from novices (ATC students) to experts (certified controllers).

**Experiment 2**

*On expertise in judgment and decision-making*

In ATC, expertise plays a crucial role in the process of information selection. Spérandio (1976) found that in radar control, the amount and type of information sought by ATCs varied according to their degree of experience and the number of aircraft: beginners did not seek the same information as experienced controllers. Another study showed that controllers’ estimates
Risk judgments in ATC

are more accurate for horizontal than for vertical separation (Bouju, 1978). Controllers determine conflict between two converging aircraft chiefly by estimating their lateral separation. Beyond ATC, studies of expertise in general have revealed that experts use more relevant information than novices (Shanteau, 1992). Thus, the quality of selection, retention and reorganization processes changes with expertise.

Some authors have argued that experts are those who have developed automatic mental processes (e.g., Anderson, 1983; Dreyfus & Dreyfus, 1986), and/or who use perceptual processes to handle situations (Chase & Simon, 1973; Lesgold et al., 1988). Thus, experts may not be “calculating”, but rather using heuristics or pattern recognition to make risk estimates in the presence of uncertainty (e.g., Klein et al., 1993). In the field of ATC, Averty (1998) observed a difference between the operating modes of experienced and less experienced controllers: with regard to conflict detection and resolution, beginners estimated the separation by means of a “mental calculus” (p. 328), which was paradoxically more accurate than the perceptual processes used by more experienced controllers (Averty, 1998). Bouju (1978) showed that in the case of converging aircraft, estimating the separation implies perceptual and logical processes, but experts rarely make any actual calculations. Perceptual processes are particularly advantageous in degraded situations, where time pressure impairs conscious cognitive processes. They reduce the workload of calculating aircraft separation, thereby sparing experts’ resources for coping with complex events. This picture of experts as intuitive decision-makers (Hogarth, 2001) may be slightly too simplistic, however. For example, other studies have shown that even in some difficult tasks with a strong perceptual component, such as radiology, some “super experts” (Raufaste et al., 1998) are characterized by the addition of a layer of conscious processes on top of the automatic layer. Also, in the ATC task, some important data, such as altitude, are provided
Risk judgments in ATC

in numerical form only, which favors analytical rather than intuitive processes (Hammond et al., 1987). In the model proposed here, we assumed that processing $D_t$, and—to a lesser extent—$D_{th}$ and $D_{t0}$ was mainly a conscious act. Because automatic, parallel processes are fast, while serial processes are slower (McClelland, 1979), response times would appear to be an interesting type of data to supplement the investigation of how the model’s variables contribute to risk judgments.

If the model is valid, differences in the mental processes of certified controllers and ATC students should be reflected by the proportion of variance explained by the model as well as by the pattern of contributions by the model variables. We therefore decided to conduct an additional experiment, in which young ATC students carried out the same task as certified controllers, so that we could compare their results in terms of judgments and response times.

Hypotheses

Within the group of certified controllers - the “experts” in this experiment (as compared with the novices, who were non-certified ATC students) -, the hypotheses were the same as in Experiment 1 (H1 and H2a to H2d). Novices generally find it more difficult to anticipate changes in dynamic situations (e.g., Boudes, 1997; Cellier et al., 1997) than experts do. Because computing $D_{t0}$, $D_{th}$, and $D_{tv}$ relies heavily on the capacity for anticipation, these variables might contribute less to the risk judgments made by novices than to those made by expert controllers (H3a). On the other hand, $D_{th}$, and $D_{tv}$ are directly related to the conventional rules that define a conflict. The novices might well stick more closely to the rules than the experts, lending greater importance to $D_{th}$, and $D_{tv}$ than the experts, and less importance to $D_{t0}$ (H3b).
Assuming that differences in processing might induce differences in response times (McClelland, 1979), we also decided to check how far the variables in the model predicted the time taken by participants to make their judgments.

Method

Participants. 88 ATC students (62.5% men and 37.5% women) from the École Nationale d’Aviation Civile in Toulouse, France, volunteered. They were divided into three levels of training. Level 1 novices ($N = 28$, 16 men and 12 women) were beginners with only one week of formal training. Level 2 novices ($N = 31$, 23 men and 8 women) had received six months of theoretical and practical training. Level 3 novices ($N = 29$, 16 men and 13 women) had received two years of theoretical and practical training. A fourth group was composed of volunteer certified controllers from Bordeaux Airport ($N = 36$, 28 men and 8 women). Their experience after qualification on the sector ranged from 1 to 21 years ($M = 10.75$ yrs, $SD = 6.28$ yrs). Their ages ranged from 30 to 56 ($M = 45.86$ yrs, $SD = 6.91$ yrs).

Variables, material and procedure. Expertise was manipulated by comparing certified controllers with ATC students who had never been certified. As in Experiment 1, $Dt_0$, $Dt_h$ and $Dt_v$ were manipulated separately. $Dt_0$ could take values from 11.37 to 30.94 nm. Averty (2005) had shown that low values of $Dt_0$ gave rise to judgments that were much too clearcut, we only built scenarios with $Dt_0 \geq 11.37$ nm, so as to increase difficulty. The aim was to make the cases more complex so as to augment expert-novice differences. For the same reason, we only retained scenarios where $Dt_h \leq 6$ nm (0, 3 or 6 nm). $Dt_v$ could be given the values 0, 5, 10, 15 or 20 nm. As the range of variable values was more restricted than in Experiment 1, the two sets of experimental results could not be directly compared. Materials were designed in the same way as
in Experiment 1, but with airline routes pertaining to Bordeaux Airport, the sector where all the experts were certified at the moment of the experiment.

**Analysis.** Analysis was basically the same as in Experiment 1, except that the intraindividual regressions were processed not only with risk judgments as DV but also with response times (RTs) as DV. In addition, expertise was introduced as a between-subjects factor in the ANOVAs. Comparisons of the overall fits (adjusted $R^2$) were made using the Mann-Whitney U-test. At the variable level, Dunnett post-hoc comparisons were conducted, with experts as the control group. With regard to RTs, 14 outlier measures were excluded (these were more than 3 SDs from the mean of the participant under consideration).

**Results and Discussion**

**Overall fit of the model.** Descriptive statistics are provided in Table 2. Intra-individual regressions confirmed that the model accounted for an important proportion of variance in risk judgments. On average, this proportion was greater in the experts (Figure 4 and Table 2). The difference was significant between experts and any of the novice groups (All $p_s < .002$), but there was no significant difference between the novice groups. For this reason, we blended all the novices into a single group and processed them as such in the subsequent analyses. Because the overall proportion of variance in risk judgments explained by the model remained satisfactory, (50% in experts, and still 20% in novices), we can say that H1 was verified.

**Contributions of the variables.** In novices, only $D_{tv}$ and $D_{th}$ contributed significantly to risk judgments, $t(87) = 6.97, p < .001, d = 0.74$ and $t(87) = 3.50, p < .001, d = 0.37$. Thus, hypotheses H2a and H2b were confirmed in the novices but neither H2c nor H2d. Moreover, when predicting response times using the model variables, only $D_{tv}$ and $D_{th}$ produced significant
negative weightings, $t(87) = 3.67, p < .001, d = 0.39$ and $t(87) = 2.43, p = .017, d = 0.26$. This result fits hypothesis H3b, i.e., novices would stay close to the conventional definition of a conflict.

In the experts, $Dtv$ made a large and significant contribution to risk judgments, $t(35) = 15.21, p < .001, d = 2.54$, but the other variables had negligible and non-significant effects on risk judgments ($d_s < .11$) (Figure 5). Thus, Experiment 2 did not fully replicate the results of Experiment 1 for $Dt_0$. This is actually not surprising, as the stimuli in the two experiments were different. The three variables, however, significantly predicted response times in experts: $Dtv$ had a negative weight $t(35) = 4.12, p < .001, d = 0.69$, whereas $Dth$ and $Dt_0$ had positive ones ($t(35) = 3.36, p = .002, d = 0.56$ and $t(35) = 3.84, p = .001, d = 0.64$). Higher $Dtv$ induced faster responses that were oriented toward “no conflict” in all groups. Higher $Dth$ induced slower responses in experts but this extra time was not accompanied by a significant contribution to risk judgments. One simple explanation may be that novices simply did not take $Dt_0$ into account and restricted their mental operations to processing $Dth$ and $Dtv$. This makes sense, because these variables relate to the very definition of a conflict. In contrast, response times suggest that experts may have tried to decide using a calculus that involved $Dt_0$ and $Dth$ to some extent (as was observed in Experiment 1) and did not succeed in making up their minds. The only trace of that processing was an increased response time, with increasing values of those variables. This increased response time makes sense if we assume that $Dt_0$ conveys uncertainty, as supported by the results of Experiment 1. It is more difficult to understand in the case of $Dth$. However, let us recall here that the highest value of $Dth$ was lower in Experiment 2 than in Experiment 1. This change was specifically intended to make the cases more difficult. Hence, it
is not surprising that the remaining values produced increased response times and turned out not to be clearcut enough to allow decision-making.

*Insert Figure 5 about here*

*Other results.* On average, the experts’ RTs were more than 21 seconds faster than those of the novices, $t(122) = 8.02, p < .001$ (see Table 2 for descriptive statistics). This is a classic result in the expertise literature (Ericsson, 1996; Ericsson & Smith, 1991). Nevertheless, the judgments themselves were not significantly different, $t(122) = 1.67, ns.$, which suggests that novices simply took more time to reach the same conclusions (see, e.g., Rantanen & Nunes, 2005 for similar results in ATC experiments).

**Conclusion of Experiment 2**

Experiment 2 demonstrated first and foremost that the model fits ATC experts better than novices (on average, 50% of intra-individual variance was accounted for in experts, vs. only 20% in novices). Clearly, the model captures processes that are acquired through experience. In experts, however, the patterns of contributions were not the same as they were in Experiment 1.

Second, $D_t$, was found to contribute to judgments, especially those made by experts. Novices may have found it difficult to estimate this distance as it could not be directly perceived. Computing this variable is cognitively complex because it involves coordinating anticipation on two distinct planes (vertical and horizontal). Moreover, these inferences are in two different data formats: numerical altitude values (flight levels), and analogical (speed vectors). This type of coordination may require subsymbolic processing, which is problematic for novices but easier for experts. Although both novices and experts may be able to perform the inferences needed to anticipate the time when the vertical distance between two aircraft will reach the conventional
1000 feet vertical separation, novices may not be sufficiently skilled to bring this information to bear and compute the horizontal distance at the same time.

General discussion

We presented a model of risk judgment in ATC, based on three variables that relate to objective parameters in this situation. In both experiments, the model provided a satisfactory account of the certified controllers’ risk judgments (44% of the variance in Experiment 1, 50% in Experiment 2, and up to 80% in some controllers), especially as we were applying a parsimonious 3-variable model to a complex decision-making process. Experiment 2 added the information that the model could capture strategies that are mainly acquired through experience.

Contribution of the variables in the model

In both experiments, we assumed that $D_{t0}$, $D_{th}$, and $D_{tv}$ would determine risk judgments in experts. The $D_{tv}$ variable—which coordinates horizontal and vertical dimensions—was always significant in certified controllers. This result must be considered in the light of Boag and colleagues, who recently showed that experts assess the separation between two aircraft on both horizontal and vertical planes (Boag et al., 2006).

One might wonder why, at first glance, the results from Experiment 2 did not fully replicate those of Experiment 1: $D_{t0}$ positively contributed in Experiment 1 but not in experiment 2; $D_{th}$ negatively contributed in Experiment 1 but not in experiment 2; $D_{tv}$ negatively contributed in both experiments but the absolute value appeared higher in Experiment 2 (-0.39 vs. -0.64); This was not completely unexpected, given that, for methodological reasons, the $D_{t0}$ values in Experiment 2 were selected so as to induce a high degree of doubt. The reduction in the range of values for this independent variable may have reduced its contribution to the judgment process and thus prevented us from detecting any differences. The same explanation holds true for $D_{th}$.
Because these high values were not used in Experiment 2, this variable had less of an effect and indeed lost all significance for the experts’ risk judgments. In our model, less informative values of $D_{th}$ may increase the necessity for participants to use strategies captured by $D_{tv}$, which was observed. As suggested by an anonymous reviewer, this explanation can be tested by computing the weights with a selection of the stimuli in Experiment 1 that were also used in Experiment 2. Under our explanation, such calculus should produce results for Experiment 1 closer to those of Experiment 2, which seems roughly to be the case. In the new results, the mean $\beta-D_{th}$ fell from -0.32 to -0.20 (0.04 in Experiment 2). The mean $\beta-D_{tv}$ jumped from -0.39 to -0.44 (-0.64 in Experiment 2). The mean $\beta-D_{t0}$ fell from 0.10 to 0.09 (-0.04 in Experiment 2). One could have expected larger changes for $\beta-D_{t0}$ but all changes are in the expected direction. This explanation, based on the properties of the stimuli, is also supported by analyses of response times in Experiment 2, which provide a more optimistic explanation: if one agrees that differences in response times reflect differences in cognitive processes - a common assumption in cognitive psychology -, the participants’ cognitive processes must have been significantly influenced by all three variables in the model. The fact that only one variable co-varied with risk judgments in Experiment 2 simply tells us that not all values of the variables influence decision-making.

Practical implications

For researchers developing user-friendly interfaces, our findings have clear implications: from the very outset, human operators must not be construed as a homogeneous group but rather as a collection of decision-makers who can be divided into several different categories. The proportion of variance explained by the two groups of certified controllers in Experiment 1 only differed by 5%. However, it is striking that experience in the sector correlated with the contribution of some variables in one group, and with other variables in the second group. Thus,
experience differentially affects cognitive processes. This finding corroborates other recent studies pointing to the importance of individual differences in analyzing expert activity (Cegarra & Hoc, 2006). Our results suggest that some features of a device maybe useful for some operators and useless for others. Thus, future control devices might offer a variety of tools allowing customization according to the strategies each particular controller prefers to use. As an example, for participants who rely heavily on $D_{th}$, the best strategy might be to provide features that help to compute the minimal horizontal distance between aircraft. For participants who rely more on $D_{tv}$, it might be more efficient to help them compute the minimal horizontal distance when vertical distance reaches 1000 feet. An ATC interface offering both types of tools would probably rapidly lead to observable differences in the ways controllers use the different tools. Indeed, this consideration has relevance way beyond the ATC field, given the range of cognitive processes that can be used to handle similar situations in different occupational contexts.

This research, especially the results about the differences between novices and experts, could also help to improve training programs. For example, this study showed that $D_{tv}$ made a strong contribution to experienced certified controllers’ judgments. Our results suggest that particular attention in the training program could be devoted to the means of assessing this distance and improving novices’ performance in handling these variables. This potential application of our model can be regarded as the most important of all, if we consider it in the light of a study (Durso et al., 1998b), highlighted by one of the anonymous reviewers of this article, which reported that 80% of en-route operational errors occur when one of the aircraft is changing altitude.

General Conclusion
This study investigated the suspected involvement of particular cognitive processes in conflict detection by expert and novice air traffic controllers in approach control. It yielded several clear-cut results:

1. Both experiments showed that in the majority of participants, be they ATC students or certified controllers, $Dtv$ was significant. This is an interesting result, as this variable takes vertical speed into account, whereas most ATC studies address situations where aircraft are cruising at constant altitude.

2. Overall, our model explained a substantial proportion of the variance in risk judgments (up to 50% in ATCs); On the other hand, one could expect the model to explain more variance because the task is constrained and because the three variables in the model could be seen as covering most of the information that could logically be taken into account. A key contribution of this paper resides precisely in the fact that even though those variables very well predict the judgments of some participants, other participants provided risk judgments based on something else. Indeed, the fact that the bulk of relevant information is included in the 3 variables of the model does not imply that controllers use it. For example, expert controllers may use security margins that are not contained in the data (e.g., Bisseret, 1995).

3. There was considerable variability in the proportions of intraindividual variance explained by the model: for some experts, the model explained virtually nothing, while for others the model explained up to 80% of variance. This clearly demonstrates that an individual-difference approach is required when investigating the cognitive processes of conflict judgment in ATC.

Undoubtedly, the cognitive processes involved in handling these variables may tap into a chunk of the most relevant information that controllers would be expected to look at. However, this model was not designed to directly model cognitive processes. A major criterion of quality
for a model is parsimony. Given that ours only had 3 variables, it accounted for a satisfactory proportion of variance. The heuristic value of the model derives from the fact that, having explained a substantial proportion of variance and identified a number of processing patterns, it highlights new directions for more in-depth research into the processes associated with the handling of those variables. This research will require more sophisticated experimental designs than the ones featured in the present article.
Authors’ note

Stéphanie Stankovic and Éric Raufaste: CNRS and University of Toulouse, Toulouse, France.


Send correspondence to Stéphanie Stankovic, CLLE-LTC, MDR, Université de Toulouse-II, 5 allées Machado, F-31058 Toulouse Cedex 9, France. E-mail: stankovic@univ-tlse2.fr
References


Biographies

Stéphanie Stankovic prepares a PhD in cognitive psychology at the University of Toulouse, France, under the supervision of the second author. She received her MS in cognitive psychology in 2005.

Éric Raufaste is a professor of cognitive psychology in the Department of Psychology at the University of Toulouse (France), where he received his PhD in 1999.

Philippe Averty is an engineer in ATC at the SDER (Sub-department for Studies and Applied Research, France). He received his PhD in cognitive psychology in 1998, at the University of Lyons, France.
Table 1

Mean risk judgments, explained variance and contributions of the variables in the model (Experiment 1).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>N</th>
<th>M (SE)</th>
<th>M (SE)</th>
<th>M (SE)</th>
<th>M (SE)</th>
<th>M (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minority</td>
<td>49</td>
<td>-1.22 (0.11)</td>
<td>0.41 (0.02)</td>
<td>0.18 (0.02)**</td>
<td>-0.13 (0.02)**</td>
<td>-0.52 (0.01)**</td>
</tr>
<tr>
<td>Majority</td>
<td>76</td>
<td>-0.95 (0.07)</td>
<td>0.46 (0.01)</td>
<td>0.05 (0.01)**</td>
<td>-0.44 (0.01)**</td>
<td>-0.31 (0.02)**</td>
</tr>
<tr>
<td>Overall</td>
<td>125</td>
<td>-1.06 (0.06)</td>
<td>0.44 (0.01)</td>
<td>0.10 (0.01)**</td>
<td>-0.32 (0.02)**</td>
<td>-0.39 (0.01)**</td>
</tr>
</tbody>
</table>

* p ≤ .05; ** p ≤ .001, when compared to zero, two-tailed
Table 2

*Explained variance and contributions of the variables in the model to risk judgments and response time (Experiment 2).*

<table>
<thead>
<tr>
<th>Risk judgments (on a -3.5 to +3.5 scale)</th>
<th>$Adjusted R^2$</th>
<th>$\beta - D_{t_0}$</th>
<th>$\beta - D_{t_h}$</th>
<th>$\beta - D_{t_v}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>$M$ (SE)</td>
<td>$M$ (SE)</td>
<td>$M$ (SE)</td>
<td>$M$ (SE)</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Novices Total</td>
<td>88</td>
<td>-0.88 (0.08)</td>
<td>0.20 (0.02)</td>
<td>0.04 (0.03)</td>
</tr>
<tr>
<td>Experts Total</td>
<td>36</td>
<td>-0.63 (0.13)</td>
<td>0.50 (0.05)</td>
<td>-0.04 (0.05)</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>-0.81 (0.07)</td>
<td>0.28 (0.02)</td>
<td>0.01 (0.03)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response Time (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Novices Total</td>
</tr>
<tr>
<td>Experts Total</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* $p \leq .05$; ** $p \leq .001$, when compared to zero, two-tailed.
Figure Captions

Figure 1. The variables in the model

Figure 2. Distribution of individual adjusted $R^2$ within certified controllers (Experiment 1). The labels above the bars represent the numbers of participants.

Figure 3. Contribution patterns (Experiment 1)

Figure 4. Distribution of individual adjusted $R^2$ as a function of Expertise (Experiment 2)

Figure 5. Contribution variables in novices and experts (Experiment 2)
Moment $t_0$, beginning of the scenario

Aircraft A (first)  
Aircraft B (second)

Moment $t_h$ of the closest horizontal distance

Aircraft A  
Aircraft B

Moment $t_v$ when vertical distance increases and passes 1000 feet

vertical distance = 1000 feet

Aircraft A  
Aircraft B
Vertical bars represent 95% confidence intervals

Beta values

Minority (N = 49)  Majority (N = 76)
Vertical bars represent 95% confidence intervals

-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3

Beta values

Novices (N = 88) Experts (N = 36)