Modeling Risk Judgments about Conflicts between Planes in Air Traffic Controllers: Individual Differences and Expertise

Stéphanie Stankovic and Éric Raufaste
CNRS and University of Toulouse-II
5, Allées A. Machado
31058 Toulouse Cedex, France.
{stankovic,raufaste}@univ-tlse2.fr

Philippe Averty
DGAC/DSNA/DTI/SDER
7, Avenue Édouard Belin
31400 Toulouse, France.
averty@cena.fr

ABSTRACT
Cognitive processes of conflict detection in ATC were investigated using a new functional model of the air traffic controller (Averty, 2005). The model proposes that three variables are used for making conflict judgments. In two experiments, the model described about 50% of the variance in experienced air traffic controllers (overall, N = 161) but only 26% in novices (N = 88). There was important intraindividual variance in conflict judgments. Moreover, controllers could be categorized in groups exhibiting different uses of the variables, which demonstrates that an individual difference approach is required when one wants to investigate cognitive processes of conflict judgment in ATC.

Keywords
Human modeling, Risk judgment, Expertise, Air traffic control, Conflict Detection.

INTRODUCTION
Two experiments investigated risk judgments about conflict between two planes by Air Traffic Controllers (ATCs). The first study tested the psychological validity of a recent model developed by one of us (Averty, 2005) at the SDER (Sub Department for Studies and Applied Research) [3]. This model attempted to describe elements that controllers use when judging whether the situation is risky or not. The objective was to test whether the model variables are actually used, and to detect individual differences in experts with regard to their use of these variables. The second study aimed at comparing a population of experienced ATCs and a population of ATC students, so as to study the relationship between the model and expertise: does the model capture some features shared among experts? Does it describe experts better than novices?

THE AIR TRAFFIC CONTROL TASK
Air Traffic Control (ATC) consists in making decisions so as to produce a fluid traffic while avoiding potential risks of conflict [1]. These decisions, based on complex information, involve hundreds of human lives.

The concept of “conflict” between planes.
Technically, an air proximity (“airprox”) occurs when the distance between two planes falls below a threshold, which is conventionally fixed at 3 or 5 Nautical Miles (NM) on the horizontal plane, and 1 Flight Level (FL), i.e., 1000 feet on the vertical plane in approach control. Conflict detection is the process by which controllers identify potential airproxes, by anticipating future positions of planes. This identification, aimed at prevention, would constitute the main task for ATCs [6]. A potential airprox cannot always be detected as soon as a plane arrives on the radar screen: the spontaneous evolution of planes may later generate a conflict. However, 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 taking-off (not visualized) and a low altitude flight. Depending how they anticipate the separation, controllers act in three different ways. A simple monitoring is done when controllers consider that the separation will remain above the minimal threshold. When the anticipated separation between planes that converge towards the same airport exceeds the threshold, controllers proceed with a regulation activity (e.g., regulation of planes’ speed) in order to preserve the minimal separation. Finally, when the estimated separation falls below the minimal threshold an action of conflict resolution is considered.

The data processed by ATCs
Three types of control correspond to three distinct activities (including in terms of cognitive processes). En route control handles flying over the territory above a certain altitude. Approach control handles the traffic arriving to or departing from airports. Airport control handles the traffic in the immediate airport surroundings. In approach control, because of climbing or descent moves controllers must particularly pay attention to altitude variations. To do so, they mainly use radar information. In the en route sector, controllers watch the traffic along airline routes. To do so, they first use non radar information provided by strips, and particularly, altitudes of potentially conflicting aircraft. Then they check the trajectories before finally checking the distance of separation on the horizontal plane (radar). The present study focuses on cognitive processes involved by approach control.

The ATC task is complex. The data to be processed are numerous (more than 25 for each individual flight, up to 600 in busy situations), diversified (they are not directly limited to traffic and may concern controllers’ work environment). Those parameters are dynamic (their values change over time even without any intervention), fuzzy (often, there is considerable imprecision in the data) and
uncertain (e.g., some devices give an estimate of an aircraft future position, say, at time ///, but it is impossible to know with certainty what the actual position of the plane will be at /). They can even be missing or incomplete (some flight parameters are not available to the controller, nor some piloting intentions).

COGNITIVE PROCESSES IN ATC
Relevance and Expertise

Due to the complexity of the ATC task and to the well-known effects of time pressure on decision-makers [18, 22], a central question concerns the quantity and the kind of data actually taken into account. Reduction in task complexity is done by means of a reduction in the total amount of information taken into account, with priority given to the most relevant [21]. Thus, whatever their level of experience, controllers build a mental representation of the situation to be handled. Among the data that controllers have to process, they select the most relevant, such as altitude, route, and so on [7, 19, 29].

Controllers’ ”situation awareness” is based upon the presence in working memory of elements sorted as a function of their importance [17]. Thus, Gronlund et al. (1998) found more working memory elements about the planes deemed as important than about the planes deemed as less important [19]. Expertise plays a crucial role in the selection process. Sperandio (1976) showed that in radar control the number and type of the pieces of information sought varied as a function of the number of planes and as a function of experience: experienced controllers or beginners did not seek the same information, independently of the amount of traffic to be handled [26]. As a matter of fact, works on expertise in general show that the quality of selection and retention processes changes depending on the level of expertise: even when experts use less information than novices, this information is more relevant [24, 25]. Thus, one can expect a greater proportion of relevant information in more experienced controllers. The concept of relevance as perceived by controllers might play a role in their information taking and memorization [14]. In line with some studies of learning, performance improvement can also be viewed as a consequence of an increase in the ability to distinguish relevant from less relevant information.

From Bisseret studies to the Averty model

In order to study conflict detection between two planes in ATCs, Bisseret (1981) built on Signal Detection Theory [28], SDT enables to estimate stimuli ”discriminability” and to estimate operators’ decision rule. More recently, Masalonis and Parasuraman (2003) found that Fuzzy SDT provided a better explanation of performance in conflict detection [20]. In a study contrasting expert and novice ATCs, Énard (1974) showed that novices were more accurate than their elders with regard to separation between two planes, because of the computations they performed to estimate this separation. Also, experts were more cautious than novices [15]. In her analysis of separation estimates, she found logical and perceptual patterns acquired through experience. Bouju (1978) showed that in the case of plane convergence, estimating the separation implied perceptual and logical processes but rarely referred to a calculus [9]. In addition, his study showed that controllers’ estimates were relatively better for horizontal than for vertical separation. Averty (1998) observed a difference between operating modes of experienced vs. less experienced controllers: with regard to conflict detection and solving, beginners estimate the separation by means of a ”mental calculus” (p. 328), the result of which being paradoxically more precise than perceptual heuristics used by more experienced ATCs [2]. Actually the perceptual heuristics are particularly advantageous in degraded situations, for example when time pressure alters cognitive processes. By using a perceptual heuristic, the expert ATCs would take advantage of a workload decrease in the calculus of aircraft separation to be able to cope with complex events. In contrast, beginners would tend to use more deliberate calculus.

In a recent study of en route control at constant altitude, Rantanen and Nunes found that ATCs use a lexicographic method for judging conflict risk [23]. This method first consists in a comparison of plane altitudes. If there is no potential conflict, the process is over. If the difference in altitudes falls below the 1000 feet conventional threshold, it remains to take into account horizontal distances and trajectories to estimate the minimum horizontal distance. There is conflict when the later falls below the conventional threshold of 3 or 5 NM. This strategy, however, cannot be applied to approach control because in the later case, much more complex, vertical separation is a dynamical variable. Independently of Rantanen and Nunes, Averty (2005) recently emphasized the role of the spatiotemporal information available on the interface and proposed a 3-variable model.

• The first variable called \( D_{t_0} \) corresponds to the moment \( t_0 \) (present time). \( D_{t_0} \) is the horizontal distance between the crossing and the fastest aircraft.

• The second variable called \( D_{t_h} \) corresponds to the moment \( t_h \) when the planes are closest on the horizontal plane. \( D_{t_h} \) is the minimal horizontal distance (ground projection).

• The third variable called \( D_{t_v} \) corresponds to the moment \( t_v \) when the two planes reach a growing separation of 1000 feet. \( D_{t_v} \) is the horizontal distance (ground projection) at \( t_v \). Note that \( t_h \) and \( t_v \) can indifferently precede one each other.

Rationale for choosing these variables: together, the two parameters \( D_{t_h} \) and \( D_{t_v} \) define the anticipated separation between planes, and thereby define conflict. Thus they concur to the detection process, and Averty hypothesized that they constitute the most important sources of variation. \( D_{t_0} \) is identical to the CPA (“closest point of approach”), a known variable in the ATC literature. According to Averty (2005), \( D_{t_0} \) is a factor contributing positively to errors in aircraft-position anticipation. As a consequence, he hypothesized it was also a factor of risk judgments about conflict. Averty designed not a process but a functional model. However, the variables can be analyzed in terms of cognitive resources. \( D_{t_0} \) is almost immediately given since it suffices to mentally extending the speed vectors drawn on the interface to find the planes’ crossing point. \( D_{t_0} \) is the
distance between this point and the fastest aircraft. \( D_{t_h} \) is a bit more complex because it requires mentally moving the two planes on the horizontal plane as a function of their respective speeds, and infer the moment when they will be closest on that plan. \( D_{t_v} \) is the most complex. It requires a vertical mental move of the two planes as a function of their respective vertical speeds (only the altitude information is provided, numerically). Then it requires finding the moment when the vertical separation is growing and reaching the 1000 feet threshold. Finally it requires inferring the horizontal positions of the planes at that moment, and then estimating the distance on that plan. Thus, computing \( D_{t_h} \) and \( D_{t_v} \) taps on the controller capacity to anticipate aircraft spatial positions. Anticipation is a difficult mental activity, for which an advantage of experts over novices is well documented [e.g., 8, 13]. Thus, if the Averty model correctly captures controllers’ processes, it can be predicted that its variables should contribute to risk judgments (Experiments I and II) and that the weights of \( D_{t_h} \) and \( D_{t_v} \) should increase with expertise (Experiment II).

### EXPERIMENT I

#### Hypotheses

This experiment is actually a re-analysis of data collected by Averty (2005). The purpose of the present work is to estimate the power of the model variables to predict risk judgments in expert ATCs. According to the first hypothesis, the model should explain a substantial part of the experts’ judgments about the risk of conflict. Also, \( D_{t_h} \), \( D_{t_v} \) and \( D_t \) should individually contribute significantly to risk judgments, independently of the two others.

For exploratory purpose, individual profiles were searched for in the patterns of weights. Such results could open the door to the investigation of cognitive styles, or individual strategies, in the controllers’ use of the model variables.

#### Method

**Participants.**
A total of 125 Expert controllers volunteered, 73.6% male, 26.4% female, with at least 2 years of professional experience after qualification. They worked on three sites, Lyon (\( N = 45 \)), Marseille (\( N = 45 \)), and Toulouse (\( N = 35 \)). Ages ranged from 26 to 56 (\( \text{mean} = 43.2, \text{SD} = 8.7 \)). The exact duration of their professional activity was not available.

**Variables.**
The three variables of the Averty model were manipulated by means of scenarios built from real radar recordings of aircraft trajectories. \( D_{t_h} \) could take the values 6, 10, 15, 21, 28, and 36 NM. Those values were retained so as to limit the proximity of planes to the conflict area. Given ATC rules and sector sizes, the question has no meaning beyond 40 NM. \( D_{t_v} \) (the horizontal distance at \( t_h \)) could take the values 0, 3, 6, 9, or 12 NM. The solving of those conflicts created a 12 NM separation on average (approach control) [2]. The necessity to maintain the sample size at a realistic level led to five values. \( D_t \) (the horizontal distance at \( t_v \)) could take the values 0, 5, 10, 15, or 20 NM. Those levels were chosen by taking into account the 2D projection at the moment of the minimal separation (1 FL). Each participant ran 78 trials provided by the combination of the independent variables values.

Due to the lack of information about the exact number of years of experience, and given the strong relationship between experience and age in controllers, the later information was used in lieu of the former in between-subject ANOVAs.

The main DV was the risk judgment itself, provided on a 8-point scale designed after analyzes of controllers’ cognitive processes during conflict judgment [5], once for each scenario. The left part of the scale concerned conflict judgments and the right part concerned no-conflict judgments. Each button was assigned an integer from 1 to 8. Thus lower values meant higher feelings of risk whereas higher values meant higher feelings of security.

**Material and procedure.**
The experiment lasted about 30mn. Each participant first answered to 4 familiarization trials, then 78 experimental trials in random order. For each case, the radar screen showed a couple of convergent planes on a given sector, one being of stable altitude while the other was either climbing or descending. The planes followed trajectories actually used around the airport where the data collection took place and participants could see the strips giving the type and flight plan of the planes. Speed vectors were displayed with a 3 minute temporal span. Participants were informed that their task would be to estimate the risk of conflict in no more than 1 minute and 15 seconds. While participants were reflecting, the radar screen was updated every 4 seconds. The radar screen also displayed the horizontal speed, the reference of the flight and the vertical speed. Answers were collected by means of a tactile screen laying below the radar screen.

**Analyses.**
In order to study the weight of the variables \( D_{t_h} \), \( D_{t_v} \), \( D_t \) in the risk judgment, specifically for each participant, three betas \( \beta_{-D_{t_h}}, \beta_{-D_{t_v}}, \) and \( \beta_{-D_t} \) were obtained using multiple linear regressions. Each beta represents the weight of the predictor on the risk judgment. The global fit of the model for a given participant was measured by the adjusted \( \text{R}^2 \). Absolute contributions were tested using one-sample \( t \) tests with 0 as reference value. Relative contributions were tested using multivariate ANOVAs. To report effect sizes, we used the adjusted \( \text{R}^2 \) for ANOVAs, and Cohen’s \( d \) for mean comparisons, with Cohen’s conventions [12] as reference: the effect is deemed small if \( \text{R}^2 \in [.01, .09] \) or \( d \in [.2, .5] \); medium if \( \text{R}^2 \in [.09, .25] \) or \( d \in [.5, .8] \); large if \( \text{R}^2 \geq .25 \) or \( d \geq .8 \). Finally, a cluster analysis was realized in order to identify potential strategies within our population of controllers. The classification was processed in two steps. First a hierarchical analysis showed that 2 classes structured the data. Second, we applied the \( K \)-means method in Statistica [27], asking for 2 classes. It maximizes inter-class distances and minimizes intra-class distances. Mean comparisons and a MANOVA with participants’ age as covariant were then used for interpretation.
whereas higher average), see Table 1 for descriptive statistics. On average, controllers expressed a slight feeling of "no conflict", i.e., significantly above 0 (t(124) = 2.16). In short, higher average, DT0 had a positive contribution (β = 0.56). Intra-individual regressions confirmed that the model captured a non negligible part of the risk feeling of security. To the contrary, DT0 and DT0 were positively bound, that is, higher values of those variables were accompanied with positive judgments (i.e., towards "no conflict"). In contrast higher values of DT0 were accompanied with judgments expressing feelings of risk. Thus, it could be that DT0 be used first since it requires no anticipation and constitutes nearly a "perceptual datum". The higher it is, the harder to estimate the potential of conflict in the situation. Thus, it can be interpreted as bringing a feeling of insecurity. To the contrary, DT0 and DT0 are directly bound to the probability of avoiding a conflict, by virtue of the very definition of conflict. Thus it is normal that their growing values be associated with a feeling of security. Interestingly, we found two different weighting patterns in the population of controllers. The majority pattern is characterized by a high weight of DT0, and a lower weight of DT0. The minority pattern is characterized by a high weight of DT0, and a lower weight of DT0. The two patterns share an equally low weight of DT0. In addition, the negative impact of age on DT0 and DT0 is not easy interpreting because it might be an effect of experience interpreting because it might be an effect of experience.
(e.g., more experienced controllers would develop some other strategies) but it might be a generational effect as well (e.g., younger controllers have grown up in a culture that is more based on image than their elders, and it might have changed their way of processing screen pictures.)

It might also be interesting to check whether the variables in the Avverty model genuinely capture some expert process or capture something that could be accomplished by participants with little or no controlling experience. Thus, we have been interested in complementing the first study by having young ATC students completing the same task and comparing their results with experts’.

**EXPERIMENT II**

**Hypotheses**

In dynamical situations, novices generally have more difficulty anticipating the evolution of the situation [e.g., 7, 11]. Since computing \( Dt_h \) and, even more, \( Dt_v \) builds heavily on the capacity of anticipation those variables should be used more, and hence weight more in expert than in non expert controllers. It was also expected that expert controllers would process information faster than novices.

**Method**

**Participants.**

A total of 88 ATC students (62.5% male and 37.5% female) from the École Nationale d’Aviation Civile (ENAC) in Toulouse, France, volunteered to participate in this study. They were three groups corresponding to various training levels. 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. In addition to the three student groups, a forth group of experts from the Airport of Bordeaux \((N = 36, 28 \) men and \( 8 \) women) participated in the study on a voluntary basis. Ages ranged from 30 to 56 \((mean = 45.9, SD = 6.9)\). The duration of their professional activity was not available.

**Variables, material, and procedure.**

Expertise was manipulated by means of the four groups described above. As in Experiment I, we manipulated independently the three variables \( Dt_0, Dt_h \) et \( Dt_v \). \( Dt_0 \) could take the values 15, 21, 28 or 36 NM. The first experiment had shown that, for low values of \( Dt_0 \), participants exhibited little ambiguity in their judgments. Thus, we only built 36 stimuli with \( Dt_0 \geq 15 \) NM. For the same reason, only scenarios with \( Dt_h \leq 6 \) NM, that is, 0, 3 or 6 NM were built. Finally, \( Dt_i \) could take the values 0, 5, 10, 15 or 20 NM. It is noteworthy that the range of variable values was more restrained than in Experiment I and therefore those results cannot be directly compared except for the existence of strategical patterns in expert information processing.

Materials were built in the same way as in Experiment I, yet with airline routes relevant to the Bordeaux Airport.

**Analyses.**

The analysis process was basically the same as in Experiment I. In addition, expertise was introduced in ANOVAs as a between-factor. Comparisons of the global fit was made using Mann-Whitney tests. Variances from the various Expertise groups were not homogeneous so we used Tamhane’s \( T \) for post-hoc tests. With regard to response times (RTs) 14 outlier measures were excluded (more than three SD from the mean of the participant under consideration). Given that expert and novice populations were clearly distinguished (e.g., see Table 2), we ran the classification algorithm separately for novices and experts but blending the three novice levels.

**Results**

**Global fit of the model.**

Descriptive statistics are provided in Table 2. Intraclass correlations confirmed that the model predicts a non negligible part of judgments about the risk of conflict. Figure 3 shows the distribution of explained variance shares. On average this share is higher in experts (from 22% in Level 2 novices to 60% in experts) (Figure 4), the difference was significant between experts and any of the novice groups (All \( p \leq .001 \), with Mann-Whitney tests) but there was no significant difference between the novice groups. However, as expected, experts were faster to respond than novices \((F_{(1,122)} = 64.26, p \leq .001, R^2 = .34)\).

**Contributions of the variables.**

In Level 1 novices, \( Dt_h \) had a little but significant contribution to risk judgments \( t(27) = 2.23, p = .035, d = .42 \). Other variables were non significant, although the effect of \( Dt_v \) was not negligible, \( d = .26 \). In Level 2 novices, \( Dt_v \) was marginally significant \( t(30) = 1.96, p < .06, d = 0.35 \). Other variables were non significant, although the effect of \( Dt_h \) was not negligible \( d = .25 \). In Level 3 novices, both \( Dt_0 \) and \( Dt_h \) were significant \( t(28) = 4.5, p < .001 \) and \( t(28) = 3.06, p < .005 \), with respectively large \((d = 0.82)\) and medium \((d = 0.53)\) effect sizes. \( Dt_0 \) was non significant \( d = 0.10 \). Finally, in experts, \( Dt_i \) had a significant and large contribution to risk judgments \( t(35) = 5.12, p < .001, d = 0.85 \), but other variables had negligible and non significant effects \( d < .15 \).

**Individual differences in risk judgments about conflicts.**

Descriptive statistics are provided in Table 2. Among the three variables, only \( Dt_i \) was significantly different over the various levels of expertise. \( F(3,120) = 6.01, p < .001, R^2 = .11 \). Tamhane post-hoc tests showed that experts gave higher weights to \( Dt_i \) than novices of Levels 1 \((p = .009, d = 0.60)\) and 2 \((p = .011, d = 0.59)\) but not significantly higher than Level 3 novices, even though the latter effect was non negligible \( d = 0.30 \). No significant difference could be found between novice groups. Some effects could have been statistically masked by the blending of participants using a variety of strategies. So we ran classifications, as in Experiment I, to check whether some patterns of variable use could emerged (Table 2). Two clusters emerged within novices. As one can see on Figure 5, there was a negative weight of \( Dt_i \) in the minority novices \((\tau(38) = -4.94, p < .001, d = 0.79)\) whereas in the majority the weight was positive \((\tau(48) = 14.29, p < .001, d = 2.04)\). There was a positive weight of \( Dt_0 \) in the minority novices \((\tau(38) = 3.87, p < .001, d = 0.62)\) whereas in the majority
### Table 2. Responses, explained variance and weights of the Averty model variables on risk judgments (Exp. II).

<table>
<thead>
<tr>
<th>Expertise</th>
<th>Mean Response</th>
<th>( R^2 )</th>
<th>( \beta - D_t )</th>
<th>( \beta - D_t )</th>
<th>( \beta - D_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Novices</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level I</td>
<td>28</td>
<td>5.56 (.14)</td>
<td>.25 (.04)</td>
<td>0.03 (0.05)</td>
<td>0.09 (0.04)</td>
</tr>
<tr>
<td>Level 2</td>
<td>31</td>
<td>5.16 (.12)</td>
<td>.22 (.03)</td>
<td>0.03 (0.05)</td>
<td>0.06 (0.04)</td>
</tr>
<tr>
<td>Level 3</td>
<td>29</td>
<td>5.44 (.16)</td>
<td>.30 (.04)</td>
<td>-0.03 (0.05)</td>
<td>0.13 (0.04)</td>
</tr>
<tr>
<td>Minority</td>
<td>39</td>
<td>5.21 (.11)</td>
<td>.23 (.03)</td>
<td>0.16 (0.04)</td>
<td>0.10 (0.04)</td>
</tr>
<tr>
<td>Majority</td>
<td>49</td>
<td>5.52 (.12)</td>
<td>.28 (.03)</td>
<td>-0.11 (0.03)</td>
<td>0.09 (0.03)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>88</td>
<td>5.38 (0.08)</td>
<td>.26 (.02)</td>
<td>0.01 (0.03)</td>
<td>0.09 (0.02)</td>
</tr>
<tr>
<td><strong>Experts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minority</td>
<td>15</td>
<td>5.05 (0.22)</td>
<td>.56 (.10)</td>
<td>0.75 (0.11)</td>
<td>0.37 (0.14)</td>
</tr>
<tr>
<td>Majority</td>
<td>21</td>
<td>5.18 (0.15)</td>
<td>.62 (.08)</td>
<td>-0.39 (0.09)</td>
<td>-0.32 (0.11)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>36</td>
<td>5.13 (0.13)</td>
<td>.60 (.06)</td>
<td>0.09 (0.12)</td>
<td>-0.03 (0.10)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>124</td>
<td>5.31 (0.07)</td>
<td>.36 (.03)</td>
<td>0.03 (0.04)</td>
<td>0.06 (0.03)</td>
</tr>
</tbody>
</table>

The first significant report is the fact that in the two studies which were carried out, the Averty model satisfactorily described expert conflict judgements. The second study
made it possible, among other things, to obtain precise results on the weight of the three variables of the model.

**Concerning the use of the Averty’s model variables**

In both experiments, we made the assumption that $D_{h0}$, $D_{v0}$, and $D_t$ would predict expert risk judgements on the basis of the model and results of Averty (2005). At first sight, our results could suggest that $D_{h0}$ does not contribute to risk judgements about conflicts in experts and non experts. However, let us recall that for methodological reasons, we had selected $D_{h0}$ values for which a feeling of doubt was possible ($D_{h0} > 15$ NM). Indeed, previous results [3] suggested that for small values of $D_{h0}$, participants hesitated much less. So the inclusion of these values in our second study could have masked the variability and thus could have harmed detection of differences between individuals and groups of individuals. The finding that $D_{h0}$ was no more contributive when these small values were withdrawn is extremely interesting. Indeed, it suggests that the treatment of the model variables is lexicographical in nature: one starts with the data easiest to perceive (thus easiest to treat cognitively) and checks whether it is decisive. If yes one concludes and the process is finished. If no, one passes to other variables and then $D_{h0}$ does not contribute anymore to the sequel of the process. The same reasoning applies for $D_t$ since it is known that it is decisive in its highest values (> 9 NM) [3]. To the contrary, these values were removed in our study, so this variable was not decisive anymore. As a matter of fact, $D_{t}$ was no more contributive, at least in experts. Then it suggests again a lexicographical process, participants would first consider this variable, which corroborates Averty (2005), but would almost totally ignore it once they realized that it was not decisive. Finally $D_t$ was really contributive to decision-making. Experts seemed to use significantly more the $D_t$ distance than novices. This distance is difficult to make up because it cannot be directly perceived. Using this variable is cognitively complex because its calculation requires coordinating the anticipation on two distinct dimensions. Indeed, $D_t$ is necessarily inferred from symbolic data, i.e., the numerical values of altitude (“flight levels”). Only then this assessment can be integrated into the mental representation. The latter requires a subsymbolic processing uneasy for novices but accessible to experts. In other words, both novices and experts are certainly able to carry out the inferences necessary to anticipate the moment when the vertical distance between two planes reaches the conventional vertical separation. Yet, novices would not be skilled enough to bring this information to bear and compute the horizontal distance at the same moment.

**Individual variability in judgments**

In the first experiment, we found two patterns of risk judgments about conflict in expert controllers. Similarly, in the second experiment individual variability was found under the form of two strategic patterns within experts and two within novices. Individual differences were previously found in the ATC literature with a study by Énard (1975) where experts were qualified as "pessimistic" and novices as "optimistic" [16]. This finding does not replicate in our experiment where the difference between novice and expert judgments was not significant. Furthermore, individual differences reported here obtained within the expert population and within the novice population as well.

How could we characterize those patterns? The first important difference lies in the way $D_t$ is used. The majority of both experts and novices gave it an important weight whereas the others seemed not using it at all. However, if we try to better understand the story told by those patterns we have to grasp the three variables at the same time. Let us consider the majority pattern in experts. Figure 5 shows that $D_t$ receives a high positive weight whereas the two other variables receive a negative weight. In other words, the higher $D_t$, the safer the situation is, which seems quite intuitive. To the contrary, the higher $D_{h0}$ or $D_{v0}$, the riskier is the situation. With regard to $D_{h0}$, it can be interpreted as mere prudence, that is, when planes are far from the crossing point, the situation appears more difficult to evaluate and thus gives rise to uncertainty feelings. The finding about $D_{h0}$ seems to be more counterintuitive and will deserve further investigation. Turning to the minority pattern, another story emerges: $D_t$ is apparently not used at all whereas $D_{h0}$ receives a positive weight. In other words, $D_{h0}$ correlates positively with judgments. The higher $D_{h0}$ is, the safer the situation is felt, which is quite intuitive. Interestingly, minority participants interpreted $D_{h0}$ in a radically different manner as compared to the majority group: they gave it high positive weights, which means that the distance between the aircraft and the intersection point provided those participants with a feeling of security. In novices the individual difference patterns are quite different since $D_{h0}$ is positively contributive for all novices regardless of the pattern. In the majority pattern, the two other variables behave in the same way as in experts. But in the minority pattern, $D_t$ receives a significant negative weight which, again, is quite counterintuitive.

**Conclusion**

In conclusion, this study provided some clear results:

1. Overall, the Averty model explains a substantial share of the variance of risk judgements by expert air controllers.
2. In the second study we learned that, in the majority of participants, either novices or experts, the most decisive criterion was $D_t$. This result is interesting since this variable takes vertical speed into account, which is different from studies of en route ATC where aircraft cruise at constant flight levels [e.g., 23].
3. There was considerable variability in the shares of intraindividual variance explained by the model: some experts did not conform at all to the model whereas for others the model explained 100% of the variance.
4. Since there was an important intraindividual variance that can be matched to specific judgment patterns, our results clearly demonstrate that an individual difference approach is required when one wants to investigate cognitive processes of conflict judgment in ATC. This finding corroborates other recent works [e.g., 10] pointing...
to the importance of analyzing expert activity in terms of cognitive styles.
5. Finally, about 40% of the variance remain to be explained. Thus, it will be necessary to develop a complementary approach based not only on perceptual data and quantitative analyses, but also on more qualitative investigation procedures (e.g., verbal protocol analysis).
For researchers developing tools in a "human like" approach, our findings have clear implications: human operators must be construed from the beginning not as a homogeneous group that can be characterized in a unique way, but rather as a collection of decision-making agents that can be categorized into various groups. When interacting with such agents, the system needs first to discover what category of agents the user pertains to. A relevant refinement of a "human-like approach" would be a "human-styles approach".

REFERENCES