

# Aviation Risk Perception: A Comparison Between Experts and Novices

Mary E. Thomson,<sup>1\*</sup> Dilek Önköl,<sup>2</sup> Ali Avciöglü,<sup>3</sup> and Paul Goodwin<sup>4</sup>

---

This article describes an exploratory investigation of the risk perceptions of experts and novices in relation to helicopter operations, under conditions where the participants are matched on various characteristics previously found to affect perceptions, such as demographic, gender, and background factors. The study reports considerable evidence of perceptual differences between the two participant groups (i.e., expert pilots and candidate pilots). We find that the experts' perceptions of relative risks are more veridical, in terms of their higher correlation with the true relative frequencies. A significant positive correlation between the flight hours and the contextual risk-taking tendency is also shown, leading the experienced pilots' choices toward risky alternatives in scenarios—a potential result of their overconfidence based on superior task performance. Possible explanations are offered for the findings and potential avenues for future research are identified.

---

**KEY WORDS:** Expert-novice comparisons; risk judgment; risk perception

## 1. INTRODUCTION

In the early years of helicopter aviation, the factor most likely to cause accidents was mechanically unreliable aircraft. However, the modern era has witnessed a reversal of this situation, with advanced technology now providing extremely capable machines. It now appears that the aircrew themselves are more deadly than the aircraft they fly. According to Baker *et al.* (2002), between 70% and 80% of aviation accidents can now be attributed to human error. But how are the inherent risks of flying judged by *expert* pilots,

and are these perceptions qualitatively different and more veridical than those given by novices?

Despite the importance of these questions, previous expert/novice comparisons in relation to risk perceptions are nonexistent. Some studies have compared experts with *lay people*, but even these have been relatively sparse and related to a few diverse domains, including toxicological sciences (Kraus *et al.*, 1992; Slovic *et al.*, 1985, 1995), the nuclear industry (Barke & Jenkins-Smith, 1993; Flynn *et al.*, 1993), ecological sciences (McDaniels *et al.*, 1997; Lazo *et al.*, 2000), and computer technology (Gutteling & Kutttschreuter, 1999, 2002). Although we must distinguish clearly between lay people, who are not receiving training in the domain of interest, and novices, the results and published critiques of these studies are likely to provide a useful starting point for expert/novice comparisons.

A common assertion that has emerged from the expert/lay people research is that the two groups differ substantially in how they perceive risk (cf. Wright *et al.*, 2000). For instance, Kraus *et al.* (1992) conclude

<sup>1</sup> Department of Psychology, Glasgow Caledonian University, Glasgow, Scotland, UK.

<sup>2</sup> Faculty of Business Administration, Bilkent University, Ankara, Turkey.

<sup>3</sup> Turkish Army Aviation School, Ankara, Turkey.

<sup>4</sup> School of Management, University of Bath, UK.

\* Address correspondence to Mary E. Thomson, Department of Psychology, Glasgow Caledonian University, Cowcaddens Road, Glasgow, Scotland G4 OBA, UK; M.Thomson@gcal.ac.uk.

that, unlike experts who perceive similar risks associated with chemicals and prescription drugs, lay people have negative attitudes toward chemicals but perceive prescription drugs positively. Similarly, Flynn *et al.* (1993) report that experts and lay people have radically different perceptions of radioactive waste risks. According to Fischhoff *et al.*, “even when experts and lay people have the same goals, they may be solving different problems . . .” (1982, p. 242).

However, Rowe and Wright (2001) indicate that expertise differences in risk perception may not be as clear-cut as previously suggested. Evaluating nine previous studies, which examined expert and lay perceptions of risk, these authors point to confounding social and demographic factors and methodological shortcomings that crowd the reported results. They argue that previous research has ignored factors that have been found to affect perceptions of risk, such as age, education, socioeconomic status, culture, and gender. According to Slovic (1999), such factors have been found to strongly correlate with risk judgments, and are likely to affect the judgments of both experts and lay persons. Yet, as pointed out by Rowe and Wright (2001), none of the expert *versus* nonexpert comparisons so far have attempted to match samples on these characteristics. For instance, these authors illustrate that quite often no/limited demographic details of the subjects are provided in the published work (e.g., Barke & Jenkins-Smith, 1993; Flynn *et al.*, 1993; Gutteling & Kuttschreuter, 1999), although it seems likely that the expert groups were older, contained more males, were better educated, and had a greater affiliation to the issues in question. In other studies, it is clear that potential confounding factors are present. For example, in a recent study comparing experts’ and lay persons’ perceptions of the millennium bug, conducted by Gutteling and Kuttschreuter (2002), the experts were more likely to be male (93% vs. 63% for lay persons), younger (mean of 38 years vs. 48 years for the lay persons) and more highly educated (88% vs. 41% for the lay persons). However, it must be acknowledged that researchers in such studies are always faced with a tradeoff between external and internal validity. In the case of the aforementioned study, for instance, because computer experts are typically young, highly educated men, the internal validity of the study may have been somewhat reduced if an attempt was made to match samples.

Another problem identified by Rowe and Wright (2001) is the lack of evidence that expert judgment of risk is more veridical than that of nonexperts. They point out that in the majority of previous studies ve-

racity was not evaluated in any shape or form. Recently, however, Wright *et al.* (2002) conducted a study that evaluated veracity and found that, although the experts outperformed the nonexperts on various measures, the differences were modest and both groups suffered from comparable global biases, such as overestimations arising from marginal and conditional probabilities.

Any examination of expert judgment should consider the concept of expertise itself. Experts are generally thought to possess special characteristics that enable them to perform better on relevant tasks than nonexperts. Societies rely on experts to influence and shape important policies that affect ordinary individuals’ lives. However, as pointed out by Rowe and Wright, “. . . it may be that the title of ‘expert’ is conferred on those who hold particular roles rather than the basis of the known accuracy of their judgments” (2001, p. 342). A senior position in a job environment, or being highly educated, does not guarantee expertise. The label of “expert” is typically acquired through peer consensus (Shanteau, 1992). For this label to be legitimate, performance should, ideally, be assessed objectively, but in many domains this is not practically possible. However, Bolger and Wright (1994) have constructed a framework for predicting the accuracy of decision performance. These authors have suggested that performance is likely to be good when both *ecological validity* (i.e., the relevance of the task to the professional experience of the subjects) and *learnability* (i.e., the availability of accurate, relevant, and objective data to base one’s decisions and the existence of adequate feedback about the accuracy of judgments in the domain) are high. Conversely, if either one of these constructs is low, performance is likely to be poor. Finally, Bolger and Wright have shown that, when both of the constructs are low, it would be unreasonable to expect professional judgment to be any better than that of subjects with no experience in the task domain.

As asserted by Bolger and Wright (1994), for ecological validity to be established, the experimental task should closely resemble the actual task performed by experts in the workplace. Experts should be more knowledgeable than nonexperts about the risks in question, and this depends largely on learnability within a domain, which, in turn, is influenced by the type of feedback that the expert typically receives while performing work-related tasks. If usable feedback is not available in a particular domain, the decisionmaker cannot learn from past errors and improve on judgmental performance.

It is worth noting that the situations discussed by Bolger and Wright, where feedback appears to have been beneficial (e.g., probability forecasts by weather forecasters, odds forecasting in horse racing, and interest rate prediction), differ in several important ways from many situations that will be of interest to risk specialists. In Bolger and Wright's situations it is possible for the expert to make regular estimates of risk and then to receive rapid feedback on outcomes after each estimate before making subsequent estimates. These situations are also characterized by high degrees of variability in the probabilities associated with the events in question. The probability of rain or of interest rates reaching a particular level will be constantly varying and therefore, to some extent, will diminish the learnability afforded by the feedback, or increase the time required for learning. This is especially likely to be the case if the feedback merely states the latest outcome (outcome feedback) since this outcome will be subject to noise and will not convey any information about the longer-term behavior of the system that is being assessed.

In contrast, the assessment of aviation risks (and many other risks that will be of interest to the "risk community" such as industrial accidents, terrorist incidents, and weather-related disasters) is concerned with rare events that occur at irregular intervals. In such circumstances, the provision of regular and repeated outcome feedback soon after expert judgments have been made will not be feasible. For events like these, the most practical way in which learnability is likely to be facilitated is through the detailed investigation of events when they occur and the publication of the results of these investigations. In this way, the causes of events will be fed back to the community of experts so that they can learn from the incident and use it in forming their perception of the risk associated with it. Individual expertise is likely to increase over time as one is presented with the results of more investigations. While feedback relating to these events will be far less frequent than in the situations described by Bolger and Wright, the probabilities that need to be assessed are likely to remain relatively constant for long periods. For example, there is no reason to suppose that the risk of an incident caused by an aircraft's hydraulic system failing will vary over relatively long periods of time. This is likely to enhance learnability and perhaps even compensate for a lack of frequent feedback.

Given the problems described above, from the outset, the present study attempts to ensure ecological validity of expertise in an aviation domain, a domain

where learnability is supported by detailed published studies of incidents when they occur (see below). Our work also tries to remove confounding factors to focus solely on the effects of expertise—that is, *expert helicopter pilots* and *candidate pilots* are matched on demographic characteristics, background, and gender with the main difference between the two groups being the expertise in the domain. Matching subjects in terms of age in the present circumstances is, of course, not practically possible. However, in order to minimize potential confounding effects, we only consider experts who are 40 or younger. Finally, the present study considers the veracity of expert judgment.

We aim to explore possible differences in risk perceptions and attitudes of experts and novices in relation to common flying risks and we seek answers to the following questions. Are the risk perceptions of experts and novices systematically different? Are experts' judgments of the relative risk of different events more veridical than those of novices? What are the underlying guiding factors of experts' and novices' risk perceptions in this context? Are experts more willing to take risks when compared to novices?

## 2. METHODOLOGY

### 2.1. Details of Samples

Two independent samples of participants were involved in this study. The first sample was composed of 36 experienced helicopter pilots (age range: 27–41 years (although pilots under the age of 40 were sought, one had actually turned 41 by the time he participated in the study), mean  $\cong$  34). Their mean flight experience was 2,412 hours (with a standard deviation of 951 hours). On average, they had been flying for 9.7 years (range: 5–14 years).

The second group was composed of 28 candidate helicopter pilots (age range: 22–24, mean  $\cong$  23). These candidates were attending a 52-week training program that is composed of 200 hours of flying time and 29 courses that involve an additional 1,157 hours of training. The two courses that are most directly relevant to this research are *Flight Safety and Prevention of Accidents* (21 hours in total) and *Emergencies* (28 hours in total). Both courses focus on emphasizing details of incidents that have happened in the past and also discuss how to avoid potential incidents (which have not happened in the past, but can potentially happen). Actions that should be taken if such incidents occur are also discussed. The mean flight experience of the candidates at the time

of this study was 55 hours (with a standard deviation of 39 hours). Given the considerable difference between the average flying experience of the candidates and established pilots, we believe that it is valid to regard the former group as *novices* within this domain.

All participants were Turkish males, and all had the same educational background, having attended the Turkish Military Academy and Army Aviation School. Therefore, to overcome the criticisms that Rowe and Wright (2001) directed at other expert/nonexpert comparisons, we attempted to match the comparison groups as much as was practically possible in terms of education, socioeconomic status, culture, and gender.

## 2.2. Domain Feedback

In the present context, in line with Bolger and Wright's (1994) requirements, it is reasonable to claim that the quality of expert feedback is quite high. For instance, when an incident/accident occurs, the "Accident Evaluation Committee" responsible for that particular region immediately goes to the area to evaluate/investigate what happened, to determine the potential causes, and so on. These committees consist of the officers in charge, highly experienced pilots, test pilots, and technicians. As a result of its investigation, the relevant committee issues an accident report, which contains a summary of the incident, information on experience/flight background of the pilots involved, statements from surviving pilots, and meteorological and other contextual conditions at the time of the incident. The report also provides a set of conclusions regarding the assumed causes of the incident/accident, which are given in terms of percentage breakdown of errors (e.g., pilot error = 15%, equipment error = 70%, and maintenance error = 15%). This report is sent to all the units; every pilot hears about the incident very quickly anyway. Pilots try to also evaluate the situation, discuss the report, and try to learn from the incident. In short, pilots do receive immediate feedback, followed by rigorous discussions of the incident.

In contrast, pilots in training do not receive such detailed information. Throughout their training, they receive only the information given by their instructors in various courses (technicians and instructor pilots teach these courses). Instructors try to illustrate the various causes of accidents with real-life incidents—and this is the sole information that the trainees receive.

**Table I.** Thirteen Aviation Risk Incidents

---

INSTRUCTIONS:

PLEASE RATE THE RISKINESS OF EACH OF THE 13 INCIDENTS BY ASSIGNING VALUES BETWEEN 0 (NO RISK) TO 100 (MAXIMUM RISK) FOR EACH INCIDENT BELOW:

1. Tail rotor shaft breakage	:_____
2. Main rotor blade failures	:_____
3. Transmission system failures	:_____
4. Hydraulic system failures	:_____
5. Electric system failures	:_____
6. Wire/water/ground strike at NOE (nap-of-earth) flight	:_____
7. Power loss	:_____
8. Power drop	:_____
9. Loss of tail rotor thrust	:_____
10. Accidents due to excessive usage in hover/taxi	:_____
11. Spatial disorientation	:_____
12. Lightning strike	:_____
13. Birdstrikes	:_____

---

## 2.3. Task Design and Procedure

Also in line with Bolger and Wright's (1994) suggestions, task and response details were designed in close consultation with highly experienced experts so that the items would reflect the most relevant risks that helicopter pilots face on a work-a-day basis. In particular, response mode constituted a critical variable in the elicitation of risk judgments and was given special consideration in attempts to ensure *ecological validity*.

First, the participants were required to rate the riskiness of 13 incidents (as listed in Table I), by assigning values between 0 (no risk) and 100 (maximum risk)—a familiar format for all participants. They were then asked to rate each incident on 15 characteristics, using seven-point scales, where each characteristic represented a dimension that is likely to affect risk perceptions in this context (Table II).

Finally, the participants were asked to consider five hypothetical scenarios regarding helicopter operations. For each scenario, the subjects were required to choose one of two action alternatives, where one alternative would be considered as a risk-taking option, while the remaining alternative would represent a risk-averse one (see Appendix A for the five scenarios presented). These were designed to identify the specific risk-taking tendencies of the two participant groups. Scenarios provide effective thinking tools that enable detailed consideration of plausible events (van der Heijden *et al.*, 2002; Wright, 2001). Furthermore, the use of scenarios in the present

**Table II.** Aviation Risk Characteristics

1. <i>Centrality</i> : To what extent is the incident important for flight safety?	Not at all central						Very central
Centrality	: 1	2	3	4	5	6	7
2. <i>Controllability</i> : How likely is it that the pilot can remain in control if the incident occurs?	Not at all controllable						Very controllable
Controllability	: 1	2	3	4	5	6	7
3. <i>Severity of consequences—personal</i> : How likely is it that an accident/mishap resulting from this incident would cause severe injury or death to pilots/passengers?	Not at all severe						Very severe
Severity of consequences—personal	: 1	2	3	4	5	6	7
4. <i>Severity of consequences—property</i> : How likely is it that an accident/mishap resulting from this incident would cause severe property damage?	Not at all severe						Very severe
Severity of consequences—property	: 1	2	3	4	5	6	7
5. <i>Importance of training</i> : To what extent is training an important factor in lessening the severity of the consequences of such an incident?	Not at all important						Very important
Importance of training	: 1	2	3	4	5	6	7
6. <i>Adequacy of training</i> : To what extent is training sufficient for dealing with such an incident?	Not at all adequate						Very adequate
Adequacy of training	: 1	2	3	4	5	6	7
7. <i>Importance of altitude</i> : To what extent is the altitude likely to affect the consequences of such an incident?	Not at all important						Very important
Importance of altitude	: 1	2	3	4	5	6	7
8. <i>Importance of crew coordination</i> : To what extent is crew co-ordination likely to affect the consequences of such an incident?	Not at all important						Very important
Importance of crew coordination	: 1	2	3	4	5	6	7
9. <i>Familiarity</i> : To what extent are pilots familiar with this kind of incident?	Not at all familiar						Very familiar
Familiarity	: 1	2	3	4	5	6	7
10. <i>Stress</i> : To what extent is stress likely to affect the consequences of such an incident?	Not at all effective						Very effective
Effect of stress on the result	: 1	2	3	4	5	6	7
11. <i>Fatigue</i> : To what extent is fatigue likely to affect the consequences of the incident in question?	Not at all effective						Very effective
Effect of fatigue on the result	: 1	2	3	4	5	6	7
12. <i>Level of overconfidence</i> : To what extent is level of overconfidence likely to affect the consequences of the incident?	Not at all effective						Very effective
Effect of overconfidence on the result	: 1	2	3	4	5	6	7
13. <i>Maintenance</i> : To what extent are technical or material defects likely to affect the consequences of the incident in question?	Not at all effective						Very effective
Effect of maintenance on the result	: 1	2	3	4	5	6	7
14. <i>Night flight</i> : To what extent is flying at night likely to impact upon the consequences of the incident?	Not at all effective						Very effective
Effect of night flight on the result	: 1	2	3	4	5	6	7
15. <i>Technology</i> : To what extent is the level of technology used in aircraft important for consequences of such an incident?	Not at all effective						Very effective
Effect of technology on the result	: 1	2	3	4	5	6	7

context is ecologically valid given that pilots typically evaluate incidents like these as they occur (candidate pilots perform these evaluations with their instructors, while the experienced pilots discuss them with other experienced pilots).

### 3. RESULTS AND DISCUSSION

Mean ratings for risk perceptions of experienced pilots and candidate pilots for the 13 specified incidents are presented in Table III. As displayed in this

table, statistically significant differences in risk perceptions are observed for five of the incidents, i.e., main rotor blade failures, hydraulic system failures, electric system failures, lightning strikes, and bird-strikes. In each of these incidents, candidates' riskiness ratings were consistently higher than that of the pilots. These findings suggest that, with experience, perceived risks associated with these particular incidents are reduced, while the risk perceptions concerning the remaining eight incidents, although still mostly reduced, are not as significantly affected.

Incident	Pilots	Candidates	<i>t</i>	<i>p</i> -Value
Tail rotor shaft breakage	81.3 (28.0)	74.8 (27.5)	-0.93	0.178
Main rotor blade failures	60.4 (24.5)	76.9 (30.7)	2.33	0.012
Transmission system failures	71.4 (25.0)	69.1 (17.3)	-0.44	0.332
Hydraulic system failures	49.4 (18.1)	60.2 (19.3)	2.28	0.014
Electric system failures	29.6 (22.8)	49.8 (20.2)	3.76	0.000
Wire/water/ground contact at NOE	81.0 (25.3)	82.0 (22.3)	0.17	0.432
Power loss	77.6 (28.0)	72.4 (21.1)	-0.85	0.198
Power drop	58.5 (25.2)	59.7 (21.3)	0.21	0.418
Loss of tail rotor thrust	60.7 (24.1)	63.5 (24.0)	0.46	0.323
Accidents due to excessive usage in hover & taxi	49.7 (25.9)	57.3 (21.1)	1.28	0.102
Spatial disorientation	72.1 (27.6)	69.8 (26.0)	-0.34	0.369
Lightning strike	48.5 (33.2)	68.9 (31.4)	2.52	0.007
Birdstrikes	30.3 (23.3)	46.3 (25.8)	2.56	0.006

**Table III.** Risk Perceptions of Pilots and Candidate Pilots for the Given Incidents—Mean Ratings (Standard Deviations) with 0: No Risk; 100: Maximum Risk

It is interesting to note that the experts' significantly lower perceptions of relative risk occur on four of the least frequent incidents (as given in Appendix B): birdstrikes, lightning strikes, electric system failures, and main motor blade failures, suggesting that the experts were better at identifying very low risks. However, to obtain a more general measure of veracity, a correlation analysis was conducted. This, indeed, indicates that the experts' judgments are more veridical: the mean correlation of the experts' risk perceptions for all 13 incidents with the relative frequencies of these incidents was significantly higher than that of the candidates ( $\bar{r}_{\text{experts}} = 0.29$ ,  $\bar{r}_{\text{candidates}} = -0.04$ ) ( $t_{52} = 5.23$ ,  $p < 0.001$ ). However, the mean correlation for the experts is clearly still quite low. One possible explanation for this is that they did not perceive the risk of an incident as simply the probability of that incident occurring. Indeed, various factors have previously been found to affect risk perceptions, including the degree of *dread* the person feels toward the risk in question, the extent to which the risk is *known* to the individual, and the likely *consequences* if the risk in question should occur (Fischhoff *et al.*, 1978; Slovic, 1987). In the present circumstances, the latter factor seemed most likely to have come into

play. That is, the pilots' notion of risk might also have included some assessment of the magnitude of the adverse effects that might be associated with the incident. Two incidents may be perceived as equally probable, but if the occurrence of one is more likely to incur serious consequences than the occurrence of the other, then the first incident might be perceived as being more risky (this is analogous to the concept of "detriment" (Adams, 1995)). To investigate this, the experts' rating of each incident in response to the question "How likely is it that an accident/mishap resulting from this incident would cause injury or death to pilots or passengers?" was used to allow their assessment of consequences to be taken into account. For each expert, the partial correlation between their perceived risk for the 13 incidents and their response to this question was calculated, with the effects of relative frequency removed. The mean partial correlation was 0.43. This is significantly greater than 0 ( $z = 7.7$ ,  $p < 0.001$ ) and indicates that, for a given relative frequency, the pilots' perceived risk will tend to be higher for incidents where they consider the probability of injury or death to be greater. This suggests that they *were* perceiving risk as a compound measure involving both probability and consequences. This may

have been because they naturally perceived riskiness as a compound of probability and consequences. Alternatively, it may have been a result of the availability heuristic (Tversky & Kahneman, 1974) in that events that have, in the past, been associated with serious consequences are more salient and hence easier to recall. Informal discussions with experienced pilots pointed toward the former explanation.

Interestingly, for the candidate pilots, training methods may have created an availability bias in the candidates. This is because potential causes of accidents, even extremely rare ones, are likely to have been discussed by instructors in relatively equal measure, thus leading candidates to overestimate the frequencies of unlikely events. Therefore, it could be that experience reduces availability bias generated by training methods to a greater extent than that induced by other factors.

To further investigate the critical factors contributing to the pilots' and the candidates' risk ratings for each incident, stepwise regression analyses were conducted. Table IV provides a summary of the results for the two participant groups. The factors that are predictive of overall riskiness ratings for each of the 13 incidents, along with their regression coefficients,  $t$ -values, associated significance levels ( $p$ -values), and the resultant adjusted coefficient of determination (adjusted  $R^2$ ) values are presented. For example, it can be seen that "night flights" and "altitude" were seen by the experienced pilots as being associated with the riskiness of power loss, presumably because high altitude and night time conditions would reduce the crew's ability to deal with a power loss incident. It should be noted that Table IV provides only those models with no multicollinearity (i.e., no overlapping of information among the predictor variables so that each listed predictor variable contributes unique information to the prediction of the relevant incident) that are still statistically significant. As can be gleaned from this table, different factors appear to influence the riskiness ratings of the pilots and the candidates for the incidents in general. The only exception is *excessive usage in hover and taxi*, which emerges as the only incident in which none of the given factors provide sufficiently unique predictive information for risk assessments of either the experienced pilots or the pilots in training. In the remaining incidents, candidate pilots appear to generally highlight human factors such as stress and crew coordination, while the experienced pilots' predictive factors revolve around task-oriented elements such as altitude, night flight, centrality, and controllability.

For example, for a hydraulic system failure, the experienced pilots' risk perceptions were influenced by the characteristics "maintenance" and "night flight," while the novices' perceptions were influenced by "crew coordination." Similarly, for a loss of tail rotor thrust incident, "night flight" was a significant characteristic for experienced pilots while for novices the significant factor was "familiarity."

Finally, the participants' answers to the scenarios were used to identify their contextual risk-taking tendencies. In particular, an "Index of Risk-Taking Tendency" (IRT) was computed as follows:

$$\text{IRT} = [(\sum p_i)/5] * 100,$$

where  $i = 1, 2, 3, 4, 5$  for the five scenarios presented to the participants, and  $p_i$  is the points received for the selected alternative in scenario  $i$ . Specifically,  $p_i = 0$  if alternative  $a$  (the riskless option) is selected in the  $i$ th scenario;  $p_i = 1$  if alternative  $b$  (the risky option) is chosen in the  $i$ th scenario. Thus, the IRT takes on values between 0 and 100, with higher IRT scores conveying a higher risk-taking tendency within the given domain. For the total participant pool (i.e., experienced pilots and candidate pilots combined), the correlation of IRT with flight hours was found to be statistically significant and positive ( $r = 0.424$ ;  $p < 0.001$ ). This result indicates that as the flight hours increase, the participant's tendency to choose the risky options increases as well. That is, experience appears to guide the choice toward the risky alternatives. However, since potential age effects could only be minimized in the present circumstances, the possibility that experience is confounded by age cannot be discounted. Nevertheless, academic research on aviation-related performance strongly suggests that this is not the case. For instance, in a study conducted by Baker *et al.* (2002) involving 3,306 pilots, crash circumstances or the prevalence of pilot errors did not significantly vary as age increased from the 40s to the 50s and the early 60s. Given this evidence, it seems rather unlikely that age affected the risk-taking tendencies of our experts.

#### 4. CONCLUSION

This article explores the risk perceptions of experts and novices in relation to helicopter operations. Specifically, we aimed to determine whether systematic differences exist between expert pilots and candidate pilots under conditions where the participants are matched on various characteristics that have been

**Table IV.** Stepwise Regression Results—Factors Predicting the Pilots' and Candidates' Risk Ratings for Each Incident

Incident	Pilots			Candidates		
	Predictor	$\beta$	$t$	Predictor	$\beta$	$t$
Tail rotor shaft breakage		none		Importance of training	-9.60	-2.74 ( $p = 0.011$ )
				Constant	81.50	
					none	
Main rotor blade failures	Consequences to property	6.60	2.57 ( $p = 0.015$ )			
	Constant	26.93				
Transmission system failures	Centrality	9.40	2.42 ( $p = 0.021$ )			
	Constant	11.48				
Hydraulic system failures	Maintenance	-9.60	-3.94 ( $p < 0.001$ )	Crew coordination	8.30	2.91 ( $p = 0.007$ )
	Night flight	4.00	2.34 ( $p = 0.026$ )	Constant	8.03	
	Constant	86.14				
Electric system failures	Centrality	7.20	3.87 ( $p < 0.001$ )	Night flight	8.20	3.01 ( $p = 0.006$ )
	Constant	4.32		Constant	4.10	
Wire/water/ground contact at NOE	Centrality	6.50	2.81 ( $p = 0.008$ )	Stress	7.80	0.21 ( $p = 0.036$ )
	Constant	39.88		Constant	35.97	
Power loss	Altitude	40.70	4.19 ( $p < 0.001$ )	Stress	8.50	3.37 ( $p = 0.002$ )
	Night flight	10.30	2.27 ( $p = 0.030$ )	Controllability	-6.60	-2.42 ( $p = 0.023$ )
	Constant	26.70		Constant	58.02	
Power drop		none		Crew coordination	6.70	2.59 ( $p = 0.016$ )
				Stress	5.00	2.17 ( $p = 0.040$ )
				Constant	-4.37	
Loss of tail rotor thrust	Night flight	5.60	2.04 ( $p = 0.050$ )	Familiarity	7.00	2.39 ( $p = 0.025$ )
	Constant	27.89		Constant	36.13	
Excessive usage in hover & taxi		none			none	
Spatial disorientation	Overconfidence	11.30	2.61 ( $p = 0.013$ )	Centrality	21.40	3.88 ( $p = 0.001$ )
	Constant	-0.94		Crew coordination	-7.60	-2.11 ( $p = 0.045$ )
				Constant	-17.51	
Lightning strike	Controllability	-9.60	-2.85 ( $p = 0.008$ )	Crew coordination	9.50	3.83 ( $p = 0.001$ )
	Altitude	5.40	2.09 ( $p = 0.045$ )	Constant	21.97	
	Constant	59.56				
Birdstrikes	Personal consequences	8.70	3.39 ( $p = 0.002$ )	Stress	7.00	2.94 ( $p = 0.007$ )
	Controllability	-6.00	-2.19 ( $p = 0.036$ )	Constant	18.46	
	Constant	38.03				



found to influence risk perceptions. In addition, to ensure ecological validity, the participants made assessments regarding incidents that reflected their work-a-day activities.

The results revealed various systematic differences between the experienced and candidate pilots. First, we find that experts' risk perceptions are significantly lower than those of novices on 5 of 13 risk incidents. This finding is in line with the results reported in most of the previous studies where experts and lay people were compared (e.g., Kraus *et al.*, 1992; Slovic *et al.*, 1985, 1995; Barke & Jenkins-Smith, 1993; Flynn *et al.*, 1993; McDaniels *et al.*, 1997; Lazo *et al.*, 2000), but contrasts considerably with Wright *et al.*, who find that experts and nonexperts "share more similarities in their risk perceptions than dissimilarities" (2000, p. 687). These discrepancies may be accounted for, in part, by the nature of the hazards under consideration, given that Wright *et al.*'s focus reflected less extreme hazards in the U.K. oil and gas industry. Clearly, as suggested by Wright *et al.*, much more research is needed before generalizations can be reasonably drawn and future work with larger number of experts and novices will be of immense importance.

A key finding of the current study reflects the fact that the experts' perceptions of risk are more *veridical* than those of novices in terms of their higher correlation with the true relative frequencies. As pointed out by Rowe and Wright (2001), a measure of veracity has been lacking in most previous studies, and yet this is arguably more important than finding out if risk perceptions of experts or nonexperts are lower or higher *per se*. Accordingly, future investigations must attempt to focus more on the veracity factor, emphasizing the importance of response mode as a potentially critical moderator that needs to be explored in this line of research.

Relatedly, the present evidence that experts' perceptions of relative risks are more veridical provides support for Bolger and Wright's (1994) model of expert performance. This is because both ecological validity and learnability are high in the present circumstances, and when these constructs are high, Bolger and Wright's model predicts that experts are likely to outperform nonexperts. We also think that further research could be designed to explore the possibility of extending Bolger and Wright's framework so that it includes other factors that might facilitate or hinder learnability, such as the constancy of the risk that is being estimated.

Experienced pilots' superior task performance (as compared to candidate pilots) is noncontestable;

furthermore, discussions with pilots suggest that an increase in flying time is clearly perceived as significantly improving task performance. Given the significant positive correlation between the flight hours and the contextual risk-taking tendency, this could help explain the current finding that experience appears to sway the pilots' choices toward risky alternatives, a potential result of their overconfidence based on improved task performance. This finding also confirms Christensen-Szalanski *et al.*'s (1983) assertion that the differences in experts' and nonexperts' risk judgments may be due to their differing exposure to the risky events.

The present study is also instrumental in exploring the underlying factors that guide experts' and nonexperts' risk perceptions in this context. Specifically, novices tend to be guided by human factors such as stress and crew coordination, while the perceptions of experienced pilots appear to be influenced by task-orientated factors such as controllability, altitude, night flight, and centrality. Such findings have important practical implications for improving training procedures and instructor communications in general. Relatedly, the systematic differences in risk perceptions of expert pilots and candidate pilots in training also have direct repercussions for designing effective systems for risk communication between different implicated parties involved in critical operations.

Given the exploratory nature of the current study, we believe that it is worth carrying out similar work in other areas where humans interact with complex technological systems and where safety depends on this interaction being carried out successfully. These studies would reveal whether the results that we found in a specific military domain also apply in nonmilitary human-technical systems. In particular, it would be helpful to discover the extent to which expert-novice differences in risk perceptions and attitudes are contingent on the specific nature of the system and its context. These studies might also reveal general relationships between the veracity of risk perceptions and risk attitudes. For example, is it a general rule that experts are more willing to take risks than novices? In addition, it would be useful to discover how learnability is facilitated in different domains and whether general principles can be uncovered to show how it can be enhanced. Areas that are worthy of exploration include physicians interacting with medical technologies, people operating within or controlling transport systems, operators of chemical plants and nuclear power stations, and people operating complex machinery on construction sites. The design of these studies will

need to be considered with great care and developed in close consultation with domain specialists if many of the weaknesses of the studies discussed by Bolger and Wright (1994) are to be avoided.

**APPENDIX A: FIVE HYPOTHETICAL SCENARIOS**

Please consider each of the following flight scenarios and select the action alternative (a or b) that best describes your probable reaction in such circumstances.

1. A main generator failure occurs during a flight and you decide to turn back to base. On the way to base, you receive a call and are asked if you can evacuate the casualties of a traffic accident. But performing this mission will take you far out of your route. What would be your probable action?
  - a. I would suggest that another vehicle evacuate the casualties due to failure in the helicopter.
  - b. Considering the generator failure not too risky for flight safety, I would perform the mission.
  
2. You are the pilot in a search and rescue helicopter in an earthquake region. While you are evacuating the wounded, you see four more injured people. But if you land and load them, you will exceed the maximum allowable load of around 500 pounds. At the same time, it is getting very dark. What would be your probable action?
  - a. I would not attempt a rescue that would overload the aircraft, thinking I would rather go back for them afterward.
  - b. I would attempt to evacuate the four wounded people.
  
3. You are flying around a vacation spot for training purposes. You see several young people waving their hands eagerly at you, indicating that they want you to fly to them. What would be your probable action?
  - a. I would continue to fly as planned and avoid the risk that would be caused by an NOE flight.
  - b. I would make an NOE flight to salute them and make the flight more exciting.

4. You have just completed a one-month mission period under difficult surroundings and you need to fly to a nearby city to catch a flight home. But heavy meteorological conditions are getting worse and visibility and ceiling are below limits. You have only two hours to catch the plane, and if you miss it, you will have to wait a whole week for the next one. What would be your probable action?
  - a. Although I would miss the plane, I would wait for better meteorological conditions and not take the risk.
  - b. I would attempt flying the helicopter to the nearby city to catch the plane.
  
5. Under the conditions of question 4, if you were the leader of a formation consisting of three helicopters, what would be your probable action?
  - a. I would not take the risk and wait for better conditions.
  - b. I would attempt flying to the nearby city to catch the plane.

**APPENDIX B: RELATIVE FREQUENCIES OF INCIDENTS**

	Relative Frequency (%)	Significantly Lower Risk Perceptions of Experts
Tail rotor shaft breakage	3.70	
Main motor blade failures	2.47	*
Transmission system failures	6.17	
Hydraulic system failures	7.40	*
Electric system failures	0.00	*
Wire/water/ground contact at NOE	11.10	
Power loss	3.70	
Power drop	11.10	
Loss of tail rotor thrust	9.87	
Accidents due to excessive usage in hover or taxi	7.40	
Spatial disorientation	4.94	
Lightning strike	0.00	*
Birdstrikes	0.00	*
Other previous minor incidents not included in the study	32.15	
	100.00	

*Note:* Asterisks indicate that experts' risk perceptions for these events were significantly lower than those of nonexperts (details given in Table III).

## REFERENCES

- Adams, J. (1995). *Risk*. London: UCL Press.
- Baker, S. P., Lamb, M. W., Grabowski, J. G., & Rebok, G. W. (2002). Human factors in aviation crashes involving older pilots. *Aviation, Space, and Environmental Medicine*, *73*, 134–138.
- Barke, R. P., & Jenkins-Smith, H. C. (1993). Politics and scientific expertise: Scientists, risk perception, and nuclear waste policy. *Risk Analysis*, *13*(4), 425–439.
- Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment. *Decision Support Systems*, *11*, 1–24.
- Christensen-Szalanski, J. J. J., Beck, D. E., Christensen-Szalanski, C. M., & Koepsell, T. D. (1983). Effects of expertise and experience on risk judgments. *Journal of Applied Psychology*, *68*, 278–284.
- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1982). Lay foibles and expert fables in judgments about risk. *American Statistician*, *36*, 240–255.
- Fischhoff, B., Slovic, P., Lichtenstein, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes toward technological risk and benefits. *Policy Science*, *8*, 127–152.
- Flynn, J., Slovic, P., & Mertz, C. K. (1993). Decidedly different: Expert and public views of risks from a radioactive waste repository. *Risk Analysis*, *13*(6), 643–648.
- Gutteling, J. M., & Kuttschreuter, M. (1999). The millennium bug controversy in the Netherlands? Experts' views versus public perceptions. In L. H. J. Goossens (Ed.), *Proceedings of the Ninth Annual Conference of Risk Analysis: Facing the Millennium* (pp. 489–493). Delft: Delft University Press.
- Gutteling, J. M., & Kuttschreuter, M. (2002). The role of expertise in risk communication: Laypeople's and experts' perception of the millennium bug in the Netherlands. *Journal of Risk Research*, *5*(1), 35–47.
- Kraus, N., Malmfors, T., & Slovic, P. (1992). Intuitive toxicology: Expert and lay judgments of chemical risks. *Risk Analysis*, *12*(2), 215–232.
- Lazo, J. K., Kinnell, J. C., & Fisher, A. (2000). Expert and layperson perceptions of ecosystem risk. *Risk Analysis*, *20*(2), 179–193.
- McDaniels, T. L., Axelrod, L. J., Cavanagh, N. S., & Slovic, P. (1997). Perception of ecological risk to water environments. *Risk Analysis*, *17*(3), 341–352.
- Rowe, G., & Wright, G. (2001). Differences in expert and lay judgements of risk: Myth or reality? *Risk Analysis*, *21*(2), 341–356.
- Shanteau, J. (1992). The psychology of experts: An alternative view. In G. Wright & F. Bolger (Eds.), *Expertise and Decision Support* (pp. 11–23). New York: Plenum.
- Slovic, P. (1987). Perception of risk. *Science*, *236*, 280–285.
- Slovic, P. (1999). Trust, emotion, sex, politics and science: Surveying the risk assessment-battlefield. *Risk Analysis*, *19*, 689–701.
- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1985). Characterizing perceived risk. In R. W. Kates, C. Hohenemser, & J. X. Kasperson (Eds.), *Perilous Progress: Managing of Technology* (pp. 91–125). Boulder, CO: Westview.
- Slovic, P., Malmfors, T., Krewski, D., Mertz, C. K., Neil, N., & Bartlett, S. (1995). Intuitive toxicology 11. Expert and lay judgments of chemical risks in Canada. *Risk Analysis*, *15*, 661–675.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
- van der Heijden, K., Bradfield, R., Burt, G., Cairns, G., & Wright, G. (2002). *The Sixth Sense: Accelerating Organizational Learning with Scenarios*. Chichester: John Wiley & Sons.
- Wright, G. (2001). *Strategic Decision Making: A Best Practice Blueprint*. Chichester: John Wiley & Sons.
- Wright, G., Bolger, F., & Rowe, G. (2002). An empirical test of the relative validity of expert and lay judgements of risk. *Risk Analysis*, *22*, 1107–1122.
- Wright, G., Pearman, A., & Yardley, K. (2000). Risk perception in the U.K. oil and gas production industry: Are expert loss-prevention managers' perceptions different from those of members of the public? *Risk Analysis*, *20*(5), 681–690.