"What is it doing now?" Results of a Survey into Automation Surprise[1]

Karel Hurts[2] and Robert J. de Boer[2]

[2]Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

Abstract. Automation Surprise has often been associated with aviation safety incidents, and numerous laboratory studies have been carried out into the phenomenon. However, as yet little empirical evidence has been collected on Automation Surprise in actual flight operations. Therefore, a survey was held under a representative sample of airline pilots. Respondents were asked about their recent experiences with Automation Surprise, about the accompanying circumstances, and about the consequences of these experiences. Results show that Automation Surprise is a relatively common phenomenon which occurs about once every month for the average pilot. However, it is also a rather inconsequential phenomenon. These and other findings are discussed with respect to their explanation and their implications for flight operations, pilot training, and aviation safety research.

Keywords: automation; automation surprise; survey; safety.

Introduction

During recent years, automation in commercial airliners has changed dramatically, from an aircraft with a basic autopilot to a highly sophisticated aircraft on which various flight functions are (partially) automated and the human operator is put more in a supervisory type of role. The increasing amount of cockpit automation sometimes results in a breakdown of the interaction between the pilot and the automated systems. One aspect of this breakdown is referred to as Automation Surprise, which can be described as: *"those cases where the automation does something without immediately preceding crew input related to the automation's action, and in which that automation action is inconsistent with crew expectations"* (Dekker, 2009). Serious aircraft accidents have already been associated with Automation Surprise, such as the Turkish Airlines TK1951 crash and the Air France AF447crash, both in 2009 (e.g., BEA, 2012; Dutch Safety Board, 2010).

It should be noted that the pilot need not be aware of the above-mentioned inconsistency, in which case there would not (yet) be Automation Surprise – though severe consequences may be the result in that case. Only after having become fully aware of the inconsistency, is there a chance – but not a guarantee – of reducing or eliminating it. The time course of automation surprise has previously been studied under laboratory conditions (De Boer, Heems, & Hurts 2014). Sometimes, Automation Surprise is colloquially defined as the situation where the pilot exclaims or thinks: "What is it doing now?", "Why did it do that?", and "What is it going to do next?" (Wiener, 1989).

In Figure 1, a conceptual model is proposed of how Automation Surprise (AS) may emerge in the interaction between automated cockpit systems and the pilot (adapted from Rankin, Woltjer, Field, and Woods, 2013). It also shows the way in which AS and other automation-related phenomena, such as automation bias and complacency, might be integrated in an overall conceptual framework.

[1] Paper presented at the 31st EAAP conference, held on September 22-26, 2014, Valletta, Malta.

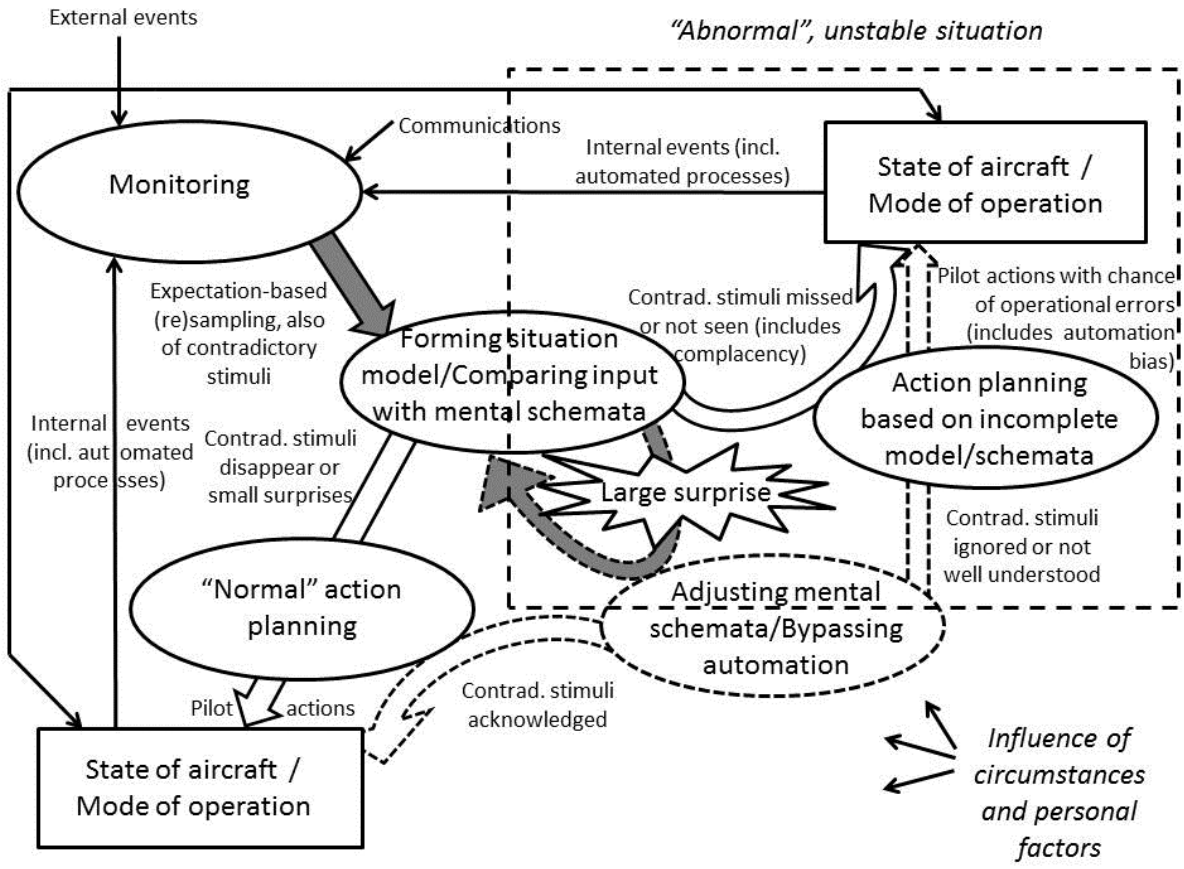


Figure 1. **Conceptual model of aircraft-pilot interaction, highlighting potential roles for automation surprise, complacency, and automation bias (adapted from Rankin, Woltjer, Field, and Woods, 2013).** Dotted lines in the bottom-right corner indicate the part of the model that is reserved for (non-routine) problem solving. Thick, dark arrows indicate monitoring/thinking activities. Thick, light arrows indicate pilot actions. Thin, black arrows indicate aircraft actions. Actions may include crew communications that feed back to the monitoring process. "Circumstances" might include degree of automation or flight phase; "personal factors" might include overall flying experience or elapsed flight duty period.

Though the precise nature of AS is still a topic of ongoing investigation, there is a fair amount of consensus regarding the following list of influencing factors:

a. *Loss of situation (mode) awareness:* situation awareness refers to "*the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future*" (Endsley, 1998). In aviation, situation awareness is considered to be less important than *mode awareness* (Harris, 2006; Oxford Aviation Academy, 2011; Sarter & Woods, 1995).
b. *Changes in pilot tasks with higher degrees of automation:* higher degrees of automation bring about qualitative changes in the tasks that remain for the pilot (Billings, 1978; Sheridan & Verplank, 1978), who is put more in a supervisory type of role. The human is not always suitable for this task (e.g., long periods of passive monitoring may cause attentional lapses or sleepiness).

c. *Inadequate design of automation interfaces.* Under high workload conditions or when abnormal or unexpected conditions arise, these interfaces may fail to inform pilots adequately (Civil Aviation Authority (UK), 2004; Harris, 2006). Attentional problems may be the result (Parasuraman & Manzey, 2010). These problems are, of course, closely related to loss of situation (mode) awareness.

The purpose of the current study was to investigate in a global way the extent to which AS occurs in practice, under what circumstances, and how the surprise was dealt with by the pilots who discovered it. It was expected that AS is not uncommon during an average flight and under average circumstances - and often can easily be undone through small, routine interventions by the pilot (Rankin, Woltjer, Field, & Woods, 2013).

Research questions and hypotheses

Supplementing and expanding existing laboratory experiments and studies with flight simulators (De Boer, Heems, & Hurts, 2014; Manzey, Reichenbach, & Onnasch, 2012), the survey approach was chosen as a practical and suitable way of querying pilots about their global experiences with AS in actual flight operations. Specifically, a questionnaire was designed to answer the following questions and test the following hypotheses:

Exploratory question/Hypothesis 1: *What is the mean prevalence* (rate of occurrence) *of AS on real flights*? We assume that AS is a relatively common phenomenon that usually is without major implications. On the basis of talks with pilots, we hypothesize that AS (a "What is it doing now?"- experience) occurs about once every month for the average pilot.

Hypothesis 2: *What are the consequences of AS? And how does AS affect trust in automation?* Because the surprise reaction may also be a sign of pilots being vigilant and trying to find out what is going on, the consequences of AS are, generally speaking, believed to be relatively mild. Similarly, trust in automation is expected to remain unchanged or decrease only slightly.

Hypothesis 3: *How does flight phase influence the experience of AS?* Based on the literature (Boeing, 2011), it is expected that difficult flight phases result in higher rates of experiencing AS than easy flight phases. This is because these flight phases are associated with the highest levels of workload, making it more likely that inconsistencies between automation behavior and pilot expectations will occur.

Hypothesis 4: *How does the degree of automation influence the experience of AS?* Despite the obvious advantages of automation, it is expected that a higher level of automation also results in problems associated with the supervisory nature of pilot tasks, thereby increasing the frequency of experiencing AS (Onnasch, Wickens, Li, & Manzey, 2013; Wiener & Curry 1980).

Hypothesis 5(a): *How does the degree of automation involved in AS influence the consequences of AS?* For the same reasons that were mentioned under Hypothesis 4, it is expected that a higher level of automation results in more severe consequences, after having experienced AS. This is despite the fact that, generally speaking, the consequences of AS are expected to be relatively mild (see Hypothesis 2).

Hypothesis 5(b): *What is the effect of degree of automation involved in AS on the amount of trust in automation?* For similar reasons, it is expected that the higher the degree of automation, the more trust in automation decreases, after an AS-experience.

Hypothesis 6: *What is the influence of flying experience on AS?* It is expected that the larger the amount of flying experience, the lower the probability of experiencing AS (Dekker & Hollnagel, 1999; Heems, De Boer, & Hurts, 2014). This is despite the fact that not all types of experience are helpful in reducing the probability of automation-related phenomena, such as complacency and automation bias (Parasuraman & Manzey, 2010).

Hypothesis 7: *What is the influence of elapsed flight duty period (FDP) on the experience of AS?* It is expected that, as a result of the building up of fatigue during a flight, the longer the elapsed FDP, the higher the probability that AS is experienced (Battelle Memorial Institute, 1998; Hartzler, 2014; Oxford Aviation Academy, 2011).

Method

Participants

Two-hundred and fifteen pilots were recruited through the Dutch Association of Commercial Pilots (VNV) and the Crew Center of KLM. In order to increase the response rate to an acceptable level, recruitment was also done through the Dutch National Aerospace Laboratory NLR and through some social networks, such as LinkedIn, PPrune, and Airwork. In order to verify that all respondents were real pilots, in the internet-based survey respondents were asked about the nature of the organization or network from which they obtained the web link containing the survey.

Equipment

Part 1 of the questionnaire: some background and demographic information was collected about the respondents, including information about gender, age (Q1), type of current and past aircraft flown, number of hours flown on current aircraft, current position, average number of flights per month (Q2), and overall flying experience in hours (Q3).

Part 2 of the questionnaire: respondents were queried about their last AS-experience or about their experiences with AS in general. Here, questions were logically grouped as follows:

a. *Last AS-experience.* How was it discovered? Briefly describe this experience (Q4). How many flights ago did this happen?(Q5) What was your flight duty period when that happened? (Q6) In what flight phase were you when that happened? (Q7) What mode was the involved system in? (Q8) What were the consequences/How was it dealt with (Q9)? Did your trust in automation change after this experience?(Q10)
b. *AS-experiences in general.* How often did AS occur in the last six months? (Q11)

Procedures

Most respondents (93%) filled out an internet-based version of the questionnaire. The remaining pilots were approached personally and filled out a paper copy of the survey. In all cases, anonymity and confidentiality of data treatment was guaranteed.

Some questions in the questionnaire were not relevant to the research question and hypotheses of this article. They pertained to the causes of AS, as perceived by the respondents, and will, therefore, not be discussed any further.

Coding of variables

Most variables appearing in the research question and hypotheses were operationalized by treating the answers as scores measured on a nominal (two or more categories) or ordinal scale. The exceptions are Q10, which consisted of a six-point rating scale, and Q4, Q5, Q8 and Q9 (open questions).

Some questions (Q1, Q2, Q3, Q6, Q11) presented a set of numerical intervals as response categories from which one was to be selected. Here, the answers were first transformed into the midscale values belonging to the intervals that had been selected, after which they were treated as scores on ratio-scale variables.

The way in which the answers to open questions Q8 and Q9 were coded will be explained in section 3. The answers to one open question (Q4) were only used for verification purposes, and were not used any further for answering or testing questions or hypotheses.

Results

General

Of the 215 returned questionnaires, 15 were discarded, because of suspicious data reliability or because only a small part of the questionnaire was filled in.

Of the remaining 200 respondents, the average age was 38 years, within a range from 23 to 58 years, stddev = 9.63 years. Moreover, 96% was male, 54% was in the rank of captain, and 42% was first officer (the balance is second officer).

The mean value for flying experience was 8867 hr, stddev = 5480 hr, with a range from 750 hr to 27500 hr. The average number of flights per month was 22.8, within a range from 3 to 43 flights, stddev = 15.09 flights.

Frequency of experiencing AS

The frequency of having AS-experiences (expressed as fraction of all flights that were AS-flights) was measured in the following two ways, using or combining the information obtained using three different survey questions:

a. The inverse of two times the answer to the open question “How many flights ago did your last AS-experience occur?” (AS-frequency score 1). This score could only be computed for those 186 (93%) respondents who had indicated to have had at least one AS-experience. Of them, 86% indicated that they had discovered the AS themselves.
b. The answer to the question “How many times did you experience AS during the last six months?” divided by six times the answer to the question “How many flights do you operate, on the average, in a month?” (AS-frequency score 2).

Note that both scores represent *relative* frequencies: the *fraction* of all flights that were AS-flights. Therefore, scores range from 0 (never) to 1 (every flight). Also note that both scores are only estimates of the “real” frequency of experiencing AS.

Some descriptive statistics are as follows:

a. *AS-frequency score 1:* 184 valid cases; median = 0.03; mean = 0.08; stddev = 0.13; skewness = 2; kurtosis = 4. Number of flights ago that last AS occurred: median = 20; mean = 71; stddev = 171.
b. *AS-frequency score 2:* 196 valid cases; median = 0.01; mean = 0.03; sttddev = 0.05; skewness = 3; kurtosis = 12. Number of AS-experiences in last 6 months: median = 1.5; mean = 2.5; stddev = 2.59.

It can be seen that lower estimations of frequency were arrived at on the basis of the number of AS-experiences in the last six months. We propose that this effect is due to the *availability effect,* known from memory psychology (Bless,1991): the pilot places greater reliance on the occurrence of AS that is more recent and more available (AS-frequency score 1). This suggests that score 1 is more reliable.

Conclusion (Exploratory question/Hypothesis 1): The mean rate of experiencing AS varied from 0.03 to 0.08 (3 to 8 out of every 100 flights), depending on the type of score used. The median rate varied from 0.01 to 0.03. These data confirm what we deduced from informal talks with pilots, i.e., that AS is a rather common phenomenon that occurs about once every month for the average pilot.

Consequences of AS: severity of consequences and change in automation trust

In order to determine the severity of the consequence of AS, the following coding procedure was used. First, the open answers given by each respondent to Q9 were classified as belonging to one of 16 types of consequence. Next, with the help of an experienced pilot, the 16 types were put in order of increasing severity (values 1 to 6). Table 1 shows the resulting frequency distribution. Note that only those respondents were considered who reported to have experienced AS at least once.

Adopting the terminology of the International Civil Aviation Organization (2013), it can be seen that the majority of the (valid) cases (98.9%) represented “incidents” that did not have (or had only mild) consequences (values 1 through 4). Values 5 and 6 (“unstable flight” and “damaged aircraft”) were mentioned in 1.1% of the cases; they may be considered “serious incidents”. Finally, “accidents”, the most severe type of consequence, was not mentioned at all by the respondents. Interestingly, these figures remind us of Heinrich’s (1931) rule: for every accident causing a major incident, there are about 29 accidents causing a minor incident, and about 300 accidents causing no injury.

Amount of change in automation trust was measured using a six-point rating scale with values ranging from 1 (“no change in trust”) to 6 (“very much change in trust”). Mirroring the results for severity of consequences, it was found that in the majority (82%) of cases automation trust had not changed - or had changed only mildly (values 1 and 2). In the remaining 18% of the cases, automation trust had changed to a larger extent (values of 3 or higher).

Conclusion (Hypothesis 2): The empirical distributions of severity of consequences and amount of change in automation trust, following an AS-experience, confirm our assumption that, though AS is a relatively common phenomenon, it usually is rather inconsequential.

Table 1. **Frequency distribution of severity of AS-consequences.** Only those 186 respondents were considered who had experienced AS at least once. Consequences are shown in order of increasing severity. Only one of the 16 consequence that are shown could be selected by each respondent.

Value	Type of consequence	N	% of total	Cumul. % of valid
1.00	NONE	32	17.2	17.5
2.00	DEGRADATION IN AUTOMATION AUTOMATION RESELECTED AFTER CORRECTION DEVIATION FMS PROCEDURE SAFETY REPORT SUBMITTED/TECH LOG	82	44.1	62.3
3.00	MANUAL TAKEOVER/CORRECTION DIVERSION GO AROUND VECTORS/HELP ATC SIGNIFICANT INCREASE IN WORKLOAD	55	29.6	92.3
4.00	LOSS OF SPEED/ALTITUDE DEVIATION ROUTE/PATH ALTITTUDE TOO LOW/HIGH SPEED TOO LOW/HIGH	12	6.5	98.9
5.00	UNSTABLE FLIGHT	1	.5	99.5
6.00	DAMAGED AIRCRAFT	1	.5	100.0
Total valid		183	98.4	
Missing		3	1.6	
Total		186	100.0	

Flight phase difficulty and the experience of AS

Results showed that, out of 11 standard flight phases (International Civil Aviation Organization, 2013), descent and approach were mentioned most frequently as being involved in an AS (27 and 31% of the respondents, respectively). Table 2 shows the breakdown of all 186 respondents who had at least one AS-experience according to the difficulty of the flight phase. Based on available aviation safety information (Boeing, 2011), the phase the pilot was in during AS was called "difficult" if take-off, climb, descent, or approach was selected as answer to question Q7. Otherwise (selection of go-around, landing, ground operations, holding, cruise-descent, cruise-climb, or cruise), the flight phase was called "easy".[2]

It can be seen that for the majority of the 165 valid cases (123 or 75%) a difficult flight phase was involved in the last AS-experience. In order to put this percentage in perspective, it can be compared to the durations of these flight phases in general, regardless of AS. Specifically, it turns out that the "difficult" flight phases of our study account, on the average,

[2] Note that there was no information available in the Boeing (2011)-publication regarding the difficulty of the phases "holding" and "go-around". Though these phases may, at times, also involve a high workload, they were mentioned by only 2.7% of the participants of this study.

for only 17% of the overall flight duration (Boeing, 2011). ("Climb" in our study equated with "Initial climb" in the latter publication, "Descent" with "Initial approach", and "Approach" with "Final approach".)

Using Bayes' theorem, the probability of experiencing AS, given a certain level of flight phase difficulty, can be computed as follows from the above-mentioned magnitudes, as well as from the fact that in our study the median relative frequency of experiencing AS (score 2), regardless of AS, is 0.01 (as discussed earlier):

$$prob(AS|flight\ phase\ difficulty\ level) = prob(flight\ phase\ difficulty\ level|AS) \times prob(AS) \ / \ prob(flight\ phase\ difficulty\ level)$$

It turns out that the first probability, also called the *posterior* probability in Bayesian decision making, was 0.003 for easy flight phases and 0.045 for difficult flight phase. This represents a fifteen-fold increase in the probability of experiencing AS for difficult flight phases, as compared to easy flight phases.

Table 2. **Breakdown of respondents according to difficulty of flight phase involved in last AS and degree of automation involved in last AS.** Only those 186 respondents were considered who had experienced AS at least once. Degree of automation was classified as "high" if one of the first three categories of Table 3 was selected, otherwise the degree of automation was classified as "low".

		Difficulty flight phase		Total
		Easy	Difficult	
Degree of automa-tion	High	32	92	124
	Low	10	31	41
Total		42	123	165

Conclusion (Hypothesis 3): Difficult flight phases were mentioned more frequently as being involved in AS than easy flight phases (in 75% of the cases). This information can be combined with the so-called prior probability of experiencing AS in our study (regardless of flight phase difficulty), as well as with industrial data regarding the average duration of easy and difficult flight phases (regardless of AS), in order to compute the probability of experiencing AS, given a certain level of flight phase difficulty (using Bayes' theorem). The latter probability turns out to be fifteen times larger for difficult flight phases (0.045) than for easy flight phases (0.003). Therefore, Hypothesis 3 is confirmed.

Degree of automation and the experience of AS

Degree of automation was measured using the unformatted answers given to open question Q8 ("In what mode was the involved system?"). Based on the expert opinion of an experienced flight instructor, these answers were recoded to the seven categories shown in Table 3. These categories vary from most automated (Autoland) to least automated (Manual flight) and in this sense form an ordinal variable for measuring Degree of automation.

It can be seen that the majority (75.3%) of the flying modes that were associated with the last AS-experience belonged to the first three (most automated) categories of Table 3

(Autoland, AP ON, AT ON, or FMS with single or dual guidance). We are not aware of any objective data about the distribution of the various automation modes across the range of all possible flights operated on all possible types of aircraft, irrespective of AS. Therefore, we are not sure how informative the figure of 75.3% is.

Looking at the effect of degree of automation on the frequency of experiencing AS (score 1), it turned out that the Spearman rank correlation between these two variables was not significant, $p > 0.10$ (see Table 3 for some insight into the way in which the mean scores for relative AS-frequency score 1 vary with degree of automation).

Table 3. **Frequency distribution for degree of automation and some descriptive statistics for relative AS-frequency score 1, broken down by degree of automation.** Only those 186 respondents were considered who had experienced AS at least once. Categories for degree of automation represent flying modes associated with the last AS. They range from a high degree of automation (top) to a low degree of automation (bottom).

	Degree of automation	N	% of valid	Rel. AS-freq. score 1 (mean)	Rel. AS-freq. score 1 (stddev)
Valid	AUTOLAND	3	1.8	0.03	0.04
	AP/AT ON, FMS GUIDANCE DUAL, APPR. MODE	111	66.9	0.10	0.15
	AP/AT ON, FMS GUIDANCE SINGLE (HOR./VERT.)	11	6.6	0.09	0.15
	AP/AT ON, MANUAL SELECT (HDG, VOR/LOC, VS)	31	18.7	0.05	0.07
	AP ON, AT OFF, MANUAL SELECT	1	0.6	0.01	-
	AP OFF, AT ON , FD ON MANUAL FLIGHT	1	0.6	0.03	-
	FD ON, MANUAL FLIGHT	8	4.8	0.05	0.05
	Total	166	100.0	0.09	0.14
Missing		20			
Total		186	100.0		

Finally, note that the two binary classifications shown in Table 2 (degree of automation: high or low; difficulty of flight phase: easy or difficult) were, statistically speaking, independent of each other, $p > 0.10$. In other words, the distribution of respondents across high and low degree of automation was not different for easy and difficult flight phases.

Conclusion (Hypothesis 4): The majority (75%) of the participants in this study reported the involvement of (one of) the three most automated flying modes (out of seven modes) in their last AS. However, it is not known to what extent this percentage merely reflects the distribution of the various modes across flights in general (irrespective of AS). Moreover, there was no significant correlation between the degree of automation that was associated

with the last AS, and the frequency of experiencing AS. Nor did the distribution of respondents across high and low degree of automation differ for easy and difficult flight phases. Therefore, Hypothesis 4 cannot be confirmed.

Effects of degree of automation on severity of AS-consequence and amount of change in automation trust

In order to assess the effect of degree of automation on severity of AS-consequences and automation-trust-change, Spearman rank correlations were computed (nonparametric correlations were chosen, because the first two variables were measured on an ordinal scale).

It turned out that the correlation with severity of AS-consequences was not significant, $p > 0.10$. On the other hand, the correlation with automation-trust-change was significant, though small, and in the expected direction: the *higher* the degree of automation involved in the last AS-experience, the *lower* the level of trust in automation became ($r = -0.16$, $p < 0.05$).

Interestingly, additional analyses showed that the latter correlation only held for the 123 cases in the "difficult flight phase"-category of Table 2.

Conclusion (Hypothesis 5a and b): Severity of AS-consequences was not sensitive to the degree of automation experienced during the last AS. Still, trust in automation decreased, as degree of automation increased, as was expected. This was particularly true for respondents who, at the time of their last AS, reported the involvement of a difficult flight phase. Therefore, Hypothesis 5a was not confirmed, but Hypothesis 5b was. These findings also confirm the importance of the level of workload in explanations of the consequences of AS.

Flying experience and the frequency of experiencing AS

In order to explore the influence of overall amount of flying experience on the frequency of experiencing AS, a multiple regression analysis was conducted on the frequencies of AS-experiences, with amount of flying experience as continuous predictor variable (measured in hours). AS-frequency score 2 was used as dependent variable. Amount of flying experience was measured using question Q4. Because regression analysis requires that dependent variables are more or less normally distributed, the frequency scores were first subjected to a log10-transformation. It was found that amount of flying experience predicted a significantly *higher* frequency of experiencing AS, $t(1) = 1.97$, $p = 0.05$, $R^2 = 0.14$.

A follow-up analysis showed that average number of flights per month per pilot had a negative and significant regression weight on the frequency of experiencing AS, $t(1) = -6.93$, $p < 0.01$: the more frequent the flying schedule, the less frequent the experience of AS. Moreover, the regression weight for total flying experience, holding average number of flights per month constant, is only small and nonsignificant, $p > 0.10$.

Conclusion (Hypothesis 6): Amount of flying experience was, unexpectedly, positively correlated with the frequency of experiencing AS. However, this effect could be explained in terms of more experienced pilots flying fewer flights per month. This, in turn, resulted in an increase of the frequency of experiencing AS. Hypothesis 6 could not be confirmed.

Probability of AS as a function of elapsed FDP

The effect of elapsed FDP on the probability of experiencing AS was computed indirectly. Specifically, the probability of any flight being an AS-flight given a specific value for FDP was computed as follows (based on the fact that for AS to occur in an FDP category, the total duration of the flight must be at least that long).

a. The relative frequency of cases (respondents) in each FDP-category was determined for those who indicated that they had experienced AS at least once (the latter information was based on question Q5).
b. The relative number of flights of a certain duration was determined, based on industry statistics (Schiphol Amsterdam Airport, 2014). Specifically, for each FDP-category the number of flights in 2013 was determined with a duration that was at least as large as the next FDP-category. Next, these values were expressed as fractions (relative frequencies) of the (constant) value for total number of flights in 2013.
c. The median value of AS-frequency score 2 was computed. It was seen before that this value is 0.01.

Similar to the way the posterior probability was computed for level of flight phase difficulty (using Bayes' theorem), the probability of any flight being an AS-flight, given a specific value for elapsed FDP, can be computed using the above-mentioned magnitudes.

Figure 2 shows the values that this probability takes on for each FDP-category. The values for the relative frequency with which an AS-flight in our study was associated with an FDP-category are also shown for reference purposes. It can be seen that the posterior probability increased from 0.002 for the lowest FDP-category to 0.01 (a fivefold increase) for the highest FDP-category.

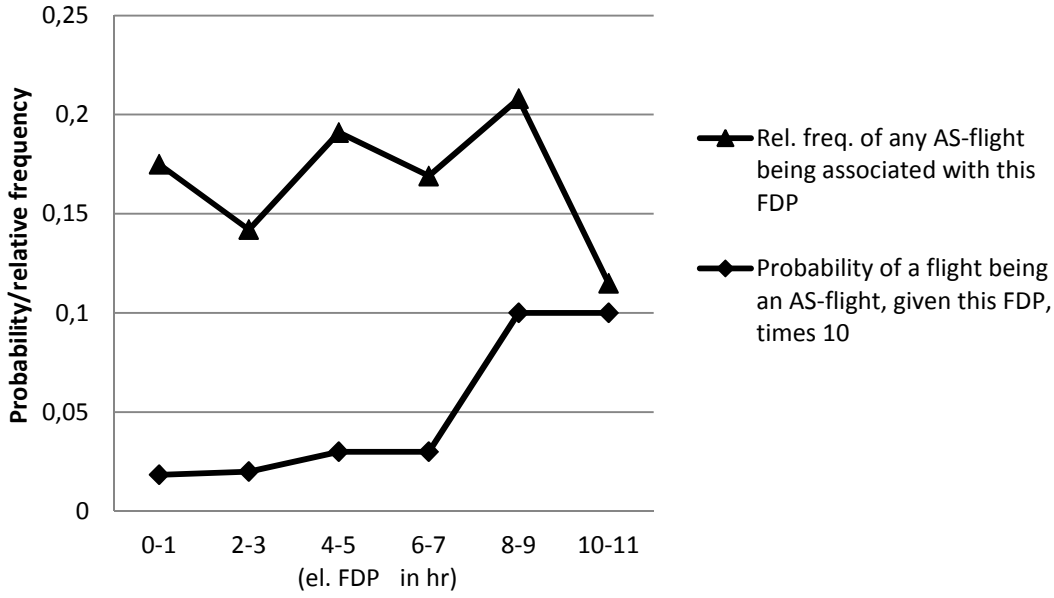


Figure 2. **Probability of a flight being an AS-flight as function of elapsed FDP.** Elapsed FDP represents amount of time (in hr) of uninterrupted flight duty at any moment during a flight.

Conclusion (Hypothesis 7): The probability of any flight being an AS-flight, given a specific value for elapsed FDP, increased with FDP and varied from 0.002 for the lowest to 0.01 for the highest FDP-category. Hypothesis 7 was confirmed.

Conclusions and Discussion

The results of this survey have expanded our knowledge about Automation Surprise (AS) in several ways. They also have implications for future aviation safety research, as well as for the future design of pilot training programs and flight operations. This will be explained below.

First, the survey approach and the focus on actual flight operations allowed us to collect and analyze a wider range of experiences regarding AS than would have been possible through a laboratory experiment or flight simulator study, though, perhaps, at a somewhat lower level of precision and reliability. AS is experienced quite regularly in actual flight operations: about once every month for an average pilot operating about 23 flights in that period. At the same time, the consequences of an average AS-experience can be considered relatively mild, the great majority of pilots (98.9%) mentioning only incidents such as manual takeover (after correction) or shifting to a lower level of automation.

Second, it was found that difficulty of the flight phase and elapsed flight duty period had positive effects on the frequency of experiencing AS. This confirmed our expectations: the more difficult the flight phase, and the higher the value of elapsed FDP, the more frequent the experience of AS. These findings deserve to be investigated further with an eye to their application in the design of personnel policies and pilot training programs.

Third, the effect of amount of flying experience on the frequency of experiencing AS was also positive, but unexpected: the more flying experience, the higher the frequency of experiencing AS. It turned out that this effect could be explained in terms of the combined presence of two other effects: (1) a negative effect of age on average number of flights per month (the older the pilot, the smaller the number of flights), and (2) a negative effect of average number of flights per month on the frequency of experiencing AS (the smaller the number of flights, the more frequently AS was experienced). In combination, these two effects suggest that opportunities to practice and rehearse flying skills get limited, once the older pilot (for personal or organizational reasons) is put on a less frequent flying schedule, thereby increasing the probability that AS is experienced.

Exploratory data analysis showed that amount of flying experience also had a (marginally significant) effect on the degree of automation that pilots thought was involved in their last AS: the *more* flying experience, the *higher* the degree of automation that was involved (Spearman rank correlation, $p < 0.10$). However, at present it is unclear how this unexpected effect should be interpreted precisely. Effects of amount of flying experience on severity of AS-consequences and on amount of automation-trust-change were not significant, $p > 0.10$. In summary, at present the precise effects of flying experience on AS-frequency and AS-consequences are not well understood and deserve further, more detailed investigations.

Fourth and finally, there was no evidence for the expectation that the degree of automation associated with the last AS has a positive effect on the frequency with which AS is experienced. Nor could the expected effect of degree of automation on the severity of AS-consequences be confirmed. In contrast, it was found that the higher the degree of automation involved in AS, the more trust in automation decreased after experiencing AS, confirming our expectation. The latter effect is notwithstanding the fact that the average amount of trust in automation hardly changed as a result of experiencing AS.

At present, it is unknown how the latter, partly conflicting and partly negative, findings should be explained or interpreted. Together with other questions about the

relationships between degree of automation and AS, they remain to be investigated more closely in future research studies.

Contact information of corresponding author

Robert J. de Boer, Amsterdam University of Applied Sciences, School of Technology, P.O. Box 1025, 1000 BA Amsterdam, The Netherlands,; email rj.de.boer@hva.nl.

Acknowledgements

The following persons are acknowledged for their contributions to this research: Roel van Leeuwen and Lévien van Kooten for their help with developing the questionnaire; Tessa Lehman de Lehnsfeld and Rieuwert Stolk for collecting the data reported in this article (all former students of Amsterdam University of Applied Sciences); and Alfred Roelen (Amsterdam University of Applied Sciences/National Aerospace Laboratory NLR) for his valuable feedback provided to several earlier versions of this article. We also wish to thank de Dutch airline pilots association VNV for their help in increasing the response rate of our survey.

References

Bless, N. (1991). Ease of retrieval as information: another look at the availability heuristic. *Journal of Personality and Social Psychology,* ***61***(2), 195–202.

Boeing (2011). *Statistical summary of commercial jet airplane accidents. Worldwide operations 1959-2010*. Seattle, WA: Dpt. Of Aviation Safety, Boeing Commercial Airplanes.

Battelle Memorial Institute (1998). *An overview of the scientific literature concerning fatigue, sleep, and the circadian cycle.* [s.l.] Battelle Memorial Institute, JIL Information Systems.

BEA (2012). *Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris.* Paris: Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile. Downloaded on August 31, 2014, from http://www.skybrary.aero/bookshelf/books/1834.pdf.

Billings, C.E. (1978). *Aviation automation: the search for a human-centered approach.* Mahwah, NJ: Erlbaum.

Civil Aviation Authority (UK) (2004). *CAA report no. 2004/10: Flight crew reliance on automation*. (Author: S. Wood, Cranfield University). Gatwick: CAA Safety Regulation Group.

De Boer, R.J., Heems, W., & Hurts, K. (2014). The duration of automation bias in a realistic setting. *The International Journal of Aviation Psychology,* ***21***(4).

Dekker, S. (2009). *Report of the flight crew human factors investigation conducted for the Dutch safety board into the accident of TK1951, Boeing 737-800 near Amsterdam Schiphol Airport, February 25, 2009*. Lund: Lund University, School of Aviation.

Dekker, S.W.A. & Hollnagel, E. (eds.) (1999). *Coping with computers in the cockpit.* Aldershot, UK: Ashgate.

Dutch Safety Board (2010). *Crashed during approach, Boeing 737-800, near Amsterdam Schiphol Airport, 25 February 2009.* The Hague: Dutch Safety Board, made available to the public on www.safetyboard.nl.

Endsley, M.R. (1998). *Situation Awareness Information Requirements for Commercial Airline Pilots.* Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting.

Harris, D. (2006). The influence of human factors on operational efficiency. *Aircraft Engineering and Aerospace Technology, **78***(1), 20–25.

Harris, D. (2011). *Human Performance on the Flight Deck*. Farnham, UK: Ashgate Publishing.

Hartzler, B.M. (2014). Fatigue on the flight deck: the consequences of sleep loss and the benefits of napping. *Accident Analysis & Prevention, **62***, 309–318.

Heinrich, H.W. (1931). *Industrial accident prevention: a scientific approach.* New York: McGraw-Hill.

International Civil Aviation Organization (2013). *Phases of flight: definition and usage notes.* Montreal: ICAO.

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: the impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*, ***6***(1), 57-87.

Onnasch, L., Wickens, C.D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: an integrated meta-analysis. *Human Factors, **56***(3), *476-488.*

Oxford Aviation Academy (2011). *Human performance & limitations*. Shoreham: Transair (UK) Ltd.

Parasuraman, R. & Manzey, D.H. (2010). Complacency and bias in human use of automation: an attentional integration. *Human Factors, **52***(3), 381-410.

Rankin, A., Woltjer, R., Field, J., & Woods, D. (2013). *"Staying ahead of the aircraft " and Managing Surprise in Modern Airliners.* Proceedings of 5th Resilience Engineering Symposium: Managing trade-offs. Soesterberg, The Netherlands.

Sarter, N. & Woods, D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory Control. *Human Factors, **37***, 5-19.

Sarter, N.B., Woods, D.D., & Billings, C.E. (1997). Automation surprises. In G. Salvendy (ed.), *Handbook of Human Factors & Ergonomics, second edition*. New York, NY: Wiley.

Schiphol Amsterdam Airport (2014). *Traffic Review 2013.*

Sheridan, T.B. & Verplank, W.L. (1978). *Human and computer control of undersea teleoperators.* (Man Machine Systems Laboratory Rep.*)* Cambridge, MA: MIT Press.

Wiener (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft.* NASA Contractor Report No. 177528. Moffett Field, CA: NASA-Ames Research Center.

Wiener, E.L. & Curry, R.E. (1980). Flight deck automation: promises and problems. *Ergonomics, **23***(10), 995-1011.