2007

# Measuring Change in Pilots' Conceptual Understandings of Autoflight

Edwin Hutchins

# MEASURING CHANGE IN PILOTS' CONCEPTUAL UNDERSTANDINGS OF AUTOFLIGHT

Edwin Hutchins
University of California San Diego
La Jolla, California

Pilots transitioning to the Airbus A320 were observed in flight and interviewed at four sample points during their first 18 months on the airplane. The interview data were analyzed by examining changes in both the relative frequencies of automation terms and the similarity of pairs of terms over time. The results show that pilots master selected modes before managed modes, and that even after 18 months of experience, their models of complex managed modes are still changing.

## Introduction

When an airline pilot transitions to a new airplane, he or she must complete a rigorous training program. If the airplane is highly automated, the pilot will receive training in the use of the autoflight system. Autoflight mode management is the process involved in understanding the character and consequences of autoflight modes, planning and selecting engagement, disengagement and transitions between modes, and anticipating automatic mode transitions made by the autoflight system itself. It has long been known that pilots are sometimes confused by the behavior of the autoflight system (Weiner, 1993; Hutchins et al. 1999; Sarter and Woods 1992, 1994). Both an industry-wide review of perceived human factors problems of flight deck automation, Funk, et al. (1999) and a special report by the FAA human factors team (FAA 1996) concluded that the complexity of automation and failures of pilot understanding of automation were thought by industry professionals to be major problems.

There is widespread agreement in the aviation industry that pilots do not acquire a complete understanding of the more advanced features of the autoflight system in training. In fact, some airlines do not attempt to teach highly automated lateral navigation (LNAV) or vertical navigation (VNAV) modes in training. It is left for the pilots to learn how to use these functions while flying on the line. Fortunately, it appears that pilots do continue to learn about the more complex functions of modern state-of-the-art airplanes long after they leave the training center. Much of what pilots know about autoflight is learned while flying in revenue service. Many pilots say it takes about 12-18 months of flying in revenue service to get comfortable with the automation. One senior Boeing 767 captain estimated that he learned approximately 60-70% of what he knows about autoflight functions while flying on the line. A typical account is that a pilot may go through three stages of automation use: In the first six months of experience the pilot is afraid of the automation and therefore makes too little use of it. In the next six months of flying the pilot gains confidence and tries to use the automation to solve every problem, thereby using the automation in inappropriate ways. Finally, the pilot understands what the automation does and what it does not do, and begins to use the automation to make the job easier only when it is appropriate to do so. While many pilots voice the beliefs contained in this progression the evidence for it is entirely anecdotal. What would a more systematic study reveal about how pilots acquire expertise with automation?

Observations from the jump seat suggest that what is learned by pilots after they leave the training center and enter revenue service includes conceptual reorganization, tuning of skills, and reassurance that what is known is sufficient to operate the airplane safely in the real flight environment. The goal of this research project was to discover how pilots' understanding of flight deck automation develops over the course of initial training and through early stages of operating experience. We hoped to document what was learned, when it was learned and how it was learned. Presumably, what pilots actually do is related to how they think about autoflight, which is in turn related to what they know about autoflight. We used primarily ethnographic methods to determine how pilots conceive of autoflight mode management (especially vertical mode management).

## Methods

In this project we chose to following a small number of pilots through the process of skill acquisition with regard to autoflight systems in the Airbus A320 passenger aircraft. To accomplish this, we recruited 15 pilots as they entered initial training with a major US-based airline and gathered data from them at regular intervals as they made their way along the initial portions of their careers flying the A320. We attempted to arrange both an interview and a jump seat observation session with each of the pilots at each sampling point. We were successful in conducting the interviews on schedule until our work was interrupted in September of 2001[1]. Due to scheduling conflicts, we were able to arrange jumpseat observations for only about half of the scheduled sample points.

**Interviews**

The interview data consist of 46 interviews with 15 pilots. Interviews were scheduled at four points in each pilot's career on the Airbus A320.

1. Initial Interview: conducted in the first few days of training. This interview sought information about the pilot's flying background and any preconceptions the pilot had about the airplane. These interviews contain discussions of attitudes toward automation in general. All initial interviews were conducted by a researcher face-to-face with the pilot, and all 15 pilots participated in an initial interview. These interviews will be referred to collectively as the Initial Interview Set (Init). Most of the subsequent interviews were conducted by a researcher[2] by telephone.
2. First line interview: conducted during the first few months of experience flying on the line. In these interviews pilots were asked to recall the most recent leg on which they were pilot-flying. These interviews will be referred to collectively as the First Line Set. (1L) One pilot was placed on medical leave, so 14 pilots participated in 1L interviews.
3. Second line interview: conducted after approximately one year of experience flying on the line. These interviews used the same format as the first line interview. (2L) One pilot was transferred back to the B737, so 13 pilots participated in 2L interviews.
4. Third line interview: conducted after approximately eighteen months of flying on the line. In these interviews, in addition to being asked to recall the most recent leg on which they were pilot-flying, the pilots were asked to describe what they would tell a pilot who is new to the airplane. (3L) Six pilots did not reach 18 months experience prior to 9/11, so 7 pilots participated in 3L interviews.

All interviews were recorded on audiotape, and transcribed by a research assistant[3] who is a pilot with knowledge of autoflight. The total interview corpus comprises approximately 336,000 words.

**Jump seat observations**

We also observed pilots from the jump seat to determine how pilots use the automation in flight. Within the constraints of the sterile cockpit rule, the jump seat provides an opportunity to talk with the pilots while they fly and it provides a rich setting for discussing things that are unclear to the pilot about autoflight functioning. Field notes from the jump seat observations complement the pilot's descriptions in the interviews. By comparing the two, we were able to confirm that the interviews were reasonably good representations of the practices the pilots actually engaged in.

## Conceptual Models of Autoflight Function

A content analysis of the interview data revealed that pilots use a small set of simple conceptual models to understand how the automation controls aircraft behavior. These basic models are known to all instrument rated pilots and are assumed by, but not generally made explicit in, airline training. Pilot models are also frequently organized around the experience of the body in the physical environment of the flight deck. Reducing thrust is typically conceptualized as "pulling", for example. This makes sense because pilots grab and pull thrust levers aft in order to manually reduce thrust. Such conceptual models are called "embodied" (Gibbs, 2006). This particular model covers not only the manual control actions of the pilot, but is also extended to the

[1] All researcher access to the flight decks of commercial airliners operating in the United States airspace was suspended before airline operations resumed a few days after the September 11, 2001 attacks. This put an immediate end to our jump seat observations. The terrorist attacks also had a profound effect on virtually all active airline pilots. The nature of the attacks and the way that airliners had been used as weapons led pilots to confront the possibility of horrifying scenarios on their own airplanes. Pilots found themselves asking, "What will I do if I get a call from the back saying that a terrorist has a knife to the throat of my lead flight attendant?" Two interviews with pilots were conducted as scheduled in the weeks following the attacks. In both cases, the participating pilots wanted, perhaps needed, to talk about the consequences of the attacks on their work. It was difficult to focus the interviews on the use of automation. These two interviews produced data that is so different from the data collected earlier in the study that we decided that it could not be used. These interviews were not transcribed. During the following months, anxieties remained high and we decided that the probability of getting usable data from additional interviews was low. We therefore ceased collecting interview data in October of 2001.

[2] The initial interviews were conducted either by Edwin Hutchins or Barbara Holder, each researcher doing about half of the initial interviews. Holder conducted a majority of the subsequent interviews, a few were conducted by Hutchins and Holder together, and one was conducted by research assistant, Howard Au. Both Holder and Au now work for Boeing.

[3] All transcription was performed by Howard Au and checked by Edwin Hutchins and Barbara Holder.

behavior of the autoflight system. Thus, the autothrust system is said to "pull the thrust back" even though the autoflight system itself does not pull anything, and when this happens in the Airbus A320 cockpit the thrust levers do not even move!

We have described the results of the content analysis elsewhere (Holder and Hutchins, 2000; Hutchins and Holder, 2001). We noted there that the number of conceptual difficulties reported concerning the descent phase of flight far outnumbers the number of difficulties reported for all other phases of flight combined. Managed descents are based on engineering principles (e.g., an energy dissipation schedule) that lie outside the realm of pilot concepts. Pilots do not normally use engineering concepts to understand autoflight. This is not surprising given that these concepts are not well represented in training materials and cannot be inferred from the behavior of the system without significant background preparation. What pilots do seem to learn on the line is when they can expect the automation to help them and how they can shape their operations to minimize automation surprises. In this paper I explore the utility of some quantitative analyses of the interview corpus as indices of change in pilots' conceptual understandings as they acquire experience in the Airbus A320.

## Quantitative Measures of Conceptual Change

To explore what simple statistical methods could reveal about changes in the pilots' conceptual structure concerning autoflight across the interviews, two types of quantitative analysis were performed: a term frequency analysis and a term co-occurrence analysis. A subset of 22 autoflight-related terms was chosen for examination. In choosing the terms, we sought a range of terms that included the most important technical terms (e.g., idle, managed), operational terms (e.g., climb, restriction) and informal pilot jargon (e.g., box).

### Term frequency analysis

Terms that occur frequently in interviews are likely to be more salient conceptually than terms that occur rarely, and changes in the relative frequencies of various terms is an indication of changing conceptual structure. We therefore computed the relative frequency of each automation term in each interview set. To ensure that we were not measuring the behavior of the interviewers, we performed the frequency analysis separately for interviewer and pilot portions of the transcripts.

*Term frequency analysis results*: The frequency analysis suggests some interesting changes in conceptual structure. For example, consider the use of the word 'computer'. This word accounts for more than a third of all autoflight related word instances in the initial interview set. The fact that the Airbus airplanes are highly computerized is THE salient fact for the pilots as they arrive at the training center. When we spoke to the pilots after they had completed training and had been flying the airplane, the use of this term dropped to about one instance in twenty of autoflight related words. Once they are on the airplane, the specifics of what the computers do become salient, and the presence of the computer is assumed rather than remarked.

The term *speed* follows a pattern of use that is almost exactly the inverse of the pattern for *computer*. *Speed* accounts for less than 5% of the instances of autoflight related words in the initial interview set. In 1L, *speed* accounts for 22% and in 2L for 26% of the autoflight related word instances. Once they are flying the airplane, the pilots' discourse concerning autoflight is dominated by the term *speed*. The frequency analysis does not reveal what it is about speed that makes it so important to the pilots. That requires a different method. The content analysis of the interviews (Hutchins and Holder, 2001) showed that pilots entering training for the airbus airplane are not yet aware of the conceptual challenges associated with the management of speed in this airplane. Pilots do not have a clear idea of what the autoflight system does in the first days of training. Once they begin flying the airplane, however, a model begins to form. It appears that in the first year on the airplane, the concepts associated with the simpler vertical modes are more salient than the concepts associated with the most highly automated vertical modes.

The following words have higher relative frequencies in the first or second line interview than they have in the last line interview: *speed*, *descent*, *climb*, *open*, *vertical*, *selected*, *select*, *restriction*. All of these terms decrease in salience in the period between one year and eighteen months on the line. The terms *open*, *selected*, and *select* are unambiguously associated with the conceptually simple "selected" guidance modes. It is widely believed in the industry that pilots use these modes most often early in their line flying. In fact, many airlines, including the one under study here, provide simulator training on the simple modes only. It is assumed that the more complex managed modes will be learned in line operations. It is known that pilots make some use of managed modes early in their line flying, but they might not discuss them at length in the interviews because they are not well understood. Like the term *speed*, *descent* is rarely mentioned in the initial interview, but peaks in the second line interview at a relative frequency of 17% (second only to *speed*). *Speed* and *vertical* are special terms because together they compose the name of one of the simple vertical guidance modes, *vertical speed*. These terms also appear in the early line interviews with high relative frequency.

The following terms have higher relative frequency in the last line interview than they have in the first two line interviews: *managed*, *mode*, *thrust*, *idle*, *autothrust*, *FMA*, *path*, *constraint*, and *target*. The increasing salience of these terms in the last line interview indicates that between twelve and 18 months on the line, the managed modes, especially, the idle thrust descent on a path defined by constraints and speed targets, become more salient concepts for the pilots. The relative frequency of the term *managed* increases with each successive phase of experience, reaching a maximum in 3L. *Mode* and *thrust* have similar profiles, but with less pronounced growth. It is probable that the third line interview has captured this learning process in progress. The fact that many of these terms, while increasing in relative frequency, still have low relative frequencies suggests that these concepts are still growing in importance.

The frequency data indicates that when pilots have a year of experience in the airplane, they talk more about the simple "selected" modes than about the more complex "managed" modes. At eighteen months, talk about the selected modes still dominates, but words that are associated with the managed modes increase in frequency.

**Term co-occurrence analysis**

The relative frequencies of terms gives us an indication of how the importance of various autoflight concepts changes with experience on the airplane, but it says nothing at all about the organization of the concepts. Co-occurrence of terms provides a simplified representation of conceptual structure. Change in conceptual organization can be tracked by representing the changing relations among terms. Two analyses of the co-occurrence of terms were performed. First, we examined each automation related term and looked at the other terms (whether related to automation or not) that tended to co-occur most frequently with that term. Second, we computed term/term similarity metrics.

To build the word-word co-occurrence matrix, a window 21 words wide is passed over the pilot conversational turns in the interview transcripts. The 'target' word is in the center of the window; the 'context' of the target word extends ten words to the right and left. The window is weighted linearly, meaning that words adjacent to the target word receive the highest co-occurrence score (in our case, 10), and those at the ends of the window receive the lowest co-occurrence score (1), with a linear progression between these two. The result is a symmetric matrix. The rows and columns are represented by all of the words used in the interviews at a particular stage of training. That is, four separate matrices are constructed: one for each of the four sets of interviews. Each cell in a matrix contains the co-occurrence value for two words. Initially, the matrix is filled with zeros. Each time two words co-occur in the same context window, their co-occurrence score is increased.

Each row (or column) in the matrix represents the co-occurrence scores of a particular word, and can be thought of as a vector in a high-dimensional space. Now consider the automation terms. We use the cosine metric to measure the angle between every pair of automation term vectors. By this measure a word will be judged semantically similar to another word not only if the two have repeatedly occurred in close proximity to one another, but also if they appeared in similar contexts: that is, if the two words occurred in close proximity to a similar set of words. As an example, consider *gem* and *jewel*. These two words tend to show up in similar contexts, and are usually substitutable for one another. As such, even though they do not frequently co-occur, they will be judged very similar by the cosine metric.

Since the cosine is a measure of relatedness, for any particular autoflight term, we can take the sum of cosines across the other autoflight terms as a measure of the centrality of the term. It turns out that the terms that increase in salience across the interview sets (as measured by relative frequency) also increase in centrality.

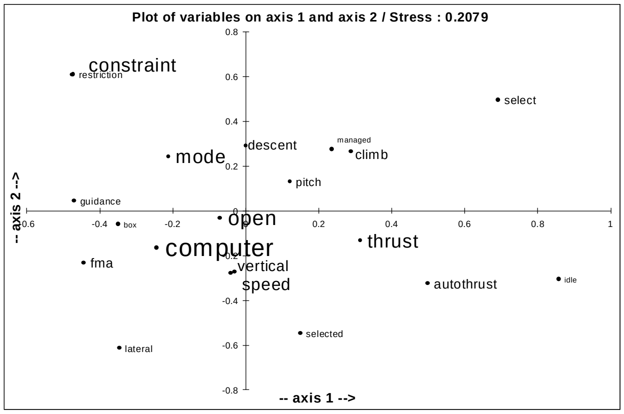


***Figure 1.*** Initial Interview (Init)

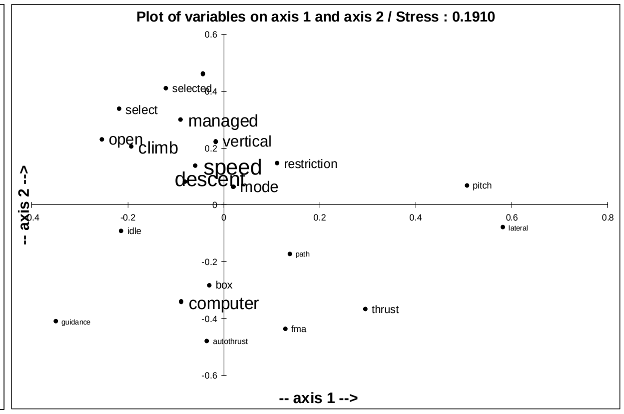


***Figure 2.*** First Line Interview (1L)

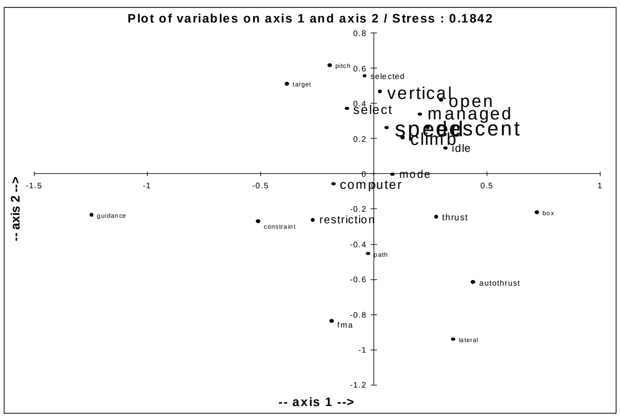


***Figure 3.*** Second Line Interview (2L)

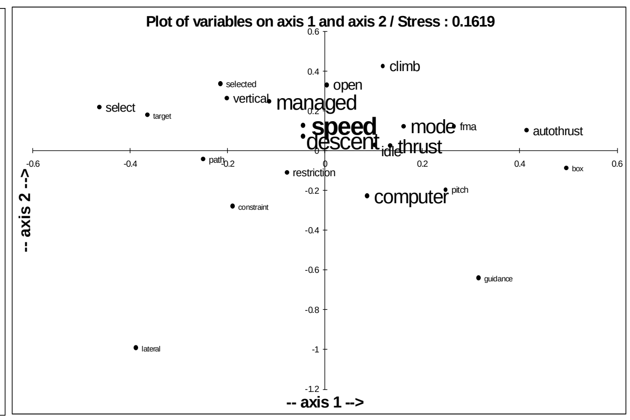


***Figure 4.*** Third Line Interview (3L)

Finally, we plot the terms in 2-dimensional space using a multidimensional scaling algorithm. Figures 1 through 4 combine these measures in a single display. Each figure represents one of the interview corpora. The terms are arranged such that the distances among terms reflect their similarities. The colors of the terms indicate their centrality as measured by sum of cosines measure for the third line interview. Warm colors indicate central terms, while cool colors indicate peripheral terms. We chose to use the centrality measure from the third line interview so that it would be easy to identify terms that are outliers in early interview data but will become central later. The font size of the term indicates the salience of the term (as measured by relative frequency in each interview set). The four composite representations in order show the emergence of an understanding of the conceptual relations of the autoflight terms.

*Term co-occurrence results* Notice the shift in color distribution across the figures. Terms that will be central in 3L (depicted in red font, e.g., *speed*, *descent*, *managed*) are spread about in the plot for the initial interview, while terms that will be peripheral in 3L (e.g., *computer* and *pitch*) are central in that plot. Coding centrality this way makes shifts in conceptual salience visually apparent.

Between the initial interview and the first line interview, *speed* has become the highest frequency term and is also the most central. Once pilots start flying the A320, they quickly discover that speed demands their attention. The simplest mode for climbs in the A320 is called open climb, and the terms *open* and *climb*, while not central, are located close together in 1L. In the first line interview, *thrust*, *path* and *idle* which will eventually be essential parts of the model are still far from central. In the second line interview *speed*, *climb*, *descent*, *open*, *managed*, and *idle* occupy central positions in the model. *Path* and *FMA* are still far from central, and *idle* is far from *thrust* and *mode*. The distinction between selected and managed modes is becoming clear, and the importance of idle thrust in descents is beginning to emerge. By the third line interview *managed*, *speed*, *descent*, *idle*, *thrust*, and *mode* have been consolidated as a central cluster. This seems to reflect an emerging understanding of the combination of features that characterize the most complex of automated modes, DES. One element of the model for DES mode is still missing, however: the computed descent path which when flown at idle thrust will produce the desired speed profile.

Two pairs of terms merit additional comments. In the first and second line interviews s*peed* and *vertical* co-occur more often than they do in the last line interview. This is evidence that the vertical speed mode is more salient in the early line interviews than it is in the last line interview. This fits the notion that early in line experience pilots are most interested in the simple modes and only gain conceptual understanding of the more complex modes after a year or more of experience. The other pair of terms is *restriction* and *constraint*. *Restriction* is an operational term used by pilots and controllers to describe elements of the flight path. *Constraint* is an engineering term, which subsumes the entities referred to by pilots as restrictions. In the Airbus pilot handbook chapter on the Flight Management System, the word *constraint* occurs 86 times, and the word *restriction* appears only once. Thus, *constraint* is the clear choice in Airbus terminology. Between the 2L and 3L interviews, the rate of use of *constraint* increases while the rate of use of *restriction* falls. For the pilots, *restriction* is more central than *constraint* at every stage, but even in the last interview set, neither term has much salience or centrality. Because constraints are essential to the definition of the path on which managed descents are flown, this fact indicates that the pilots' conceptual understandings of DES mode is still incomplete after 18 months of experience on the line.

## Discussion

The quantitative analysis provides clear evidence that a great deal of learning takes place after pilots leave training. The pilots appear to understand the basic vertical navigation modes by the time they have completed a year flying the airplane, but they are probably still in the process of acquiring an understanding of the more complex managed modes even after they have logged 18 months in the airplane. The quantitative analysis does not reveal the sources of conceptual troubles, but it does provide strong evidence that conceptual change continues for at least 18 months of line flying and it reveals the order in which conceptual elements are added to the conceptual model of the most complex modes. The quantitative analyses reveal patterns that were not apparent in the qualitative analysis. For example, while the qualitative analysis revealed that pilots used relatively simple conceptual models throughout the learning process, the quantitative date show that terms associated with the simpler autoflight modes actually peak and then decline in the first 18 months of flying experience. The term co-occurrence data show that over the same period, pilots form a complex, but still incomplete model of the most highly automated vertical navigation mode, DES.


## Acknowledgements

Thanks to Howard Au for transcription and coding assistance, Robert Liebscher for data analysis, and Barbara Holder for help with all aspects of the project. The work was funded by NASA AMES Research Center NAG 2-1313. Stephen Casner was the technical monitor. Most of all I thank the pilots who have so generously contributed their time and wisdom to this project.


## References


FAA Human Factors Team, (1996) The interfaces between flight crews and modern flight deck systems. FAA special report.

Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Sugoteguh, C., and Owen, G. (1999) Flight Deck Automation Issues. *International Journal of Aviation Psychology 92*: 109-124.

Gibbs, R. (2006) *Embodiment and Cognitive Science*. New York: Cambridge University Press.

Holder, B. and Hutchins, E. (2000) Conceptual models for understanding an encounter with a mountain wave. *Proceedings of HCI-Aero 2000*, Toulouse, France.

Hutchins, E. Holder, B., and Hayward, M. (1999) Pilot attitudes toward automation. UCSD Cognitive Science technical report.

Hutchins, E. and Holder, B. (2001) What Pilots Learn about Autoflight While Flying on the Line. 11th *International Symposium on Aviation Psychology*.

Sarter, N, and Woods, D. (1992) Pilot interaction with cockpit automation: operational experiences with the flight management system. *International Journal of Aviation Psychology 24*:303-322.

Sarter, N, and Woods, D. (1994) Pilot interaction with cockpit automation II: An experimental study of pilots' models and awareness of the flight management system. *International Journal of Aviation Psychology 41*:1-28.

Wiener, E. (1993) Crew coordination and training in the advanced-technology cockpit. In *Cockpit Resource Management* E. Wiener, B. Kanki, and R. Helmreich, Eds., Academic Press.