

**JATE**

Journal of Aviation Technology and Engineering 3:2 (2014) 20–27

## An Examination of Long-Term Working Memory Capacity

Randall L. Triplett, Joseph M. Jaworski, and Kelly J. Neville

*Embry-Riddle Aeronautical University*

---

### Abstract

The purpose of this research was to explore experts' memory capacity and the strategies experts use to achieve that capacity. Expert air traffic controllers were asked to recall traffic information during two radar and two nonradar scenarios. During radar scenarios, air traffic information was communicated aurally and displayed visually. During nonradar scenarios, air traffic information was communicated aurally only. Qualitative recall data assessment revealed an average capacity of five to eight aircraft in radar scenarios and three to six in nonradar scenarios, with two to three details recalled for most recalled aircraft. Recalled details and order of detail recall were highly consistent across experts, which suggest that aircraft details were organized and stored within larger conceptual knowledge structures. Recall patterns were additionally suggestive of frames containing slots designated for holding specific data types, structures described by Klein's Data/Frame Model of Sensemaking. The extent of information recalled and its organization are additionally consistent with the use of long term working memory to extend working memory capacity; however, they do not rule out the use of working memory alone. Differences between radar and nonradar conditions were observed. Greater overall recall and greater and earlier recall of certain data-tag elements were observed in radar scenarios. In nonradar scenarios, greater and earlier recall of assigned actions were observed. Evaluation of experts' descriptions of their recall processes suggested primarily visuospatial information encoding in both conditions and a lack of support for differences in the use of visuospatial or verbal encoding.

**Keywords:** explore long-term working memory, air traffic control, working memory, working memory capacity

---

---

### About the Authors

**Randall L. Triplett** is a USAF retired air traffic controller. He received his BS in Professional Aeronautics from Embry-Riddle Aeronautical University and his MA in Education and Human Development from George Washington University. He has 24 years of experience as an air traffic control instructor at the colligate level, and is currently employed as an assistant professor of Air Traffic Management at Embry-Riddle Aeronautical University. He is interested in the use of simulators in training. Correspondence concerning this article should be sent to [tripl4c8@erau.edu](mailto:tripl4c8@erau.edu).

**Joseph M. Jaworski** received his BS in Air Traffic Management from Embry-Riddle Aeronautical University in 2010 and an MS in Human Factors and Systems from Embry-Riddle Aeronautical University in 2012. His research interests include psychology, human factors, and air traffic control.

**Kelly J. Neville** received a PhD in Experiential Psychology from Rice University and a BA in Psychology, Business, and Spanish from Trinity University. Currently an associate professor in the Department of Human Factors and Systems at Embry-Riddle Aeronautical University, she is interested in Human Factors in complex environments.

Working memory refers to the conceptual storage area where information is held while being actively processed. One of the most actively pursued working memory research topics has been the issue of working memory capacity – how much information can be held in working memory at one time? This question becomes especially interesting when targeted at experts—how much information can be actively held and used in working memory by experts working in their domain of expertise?

Working memory capacity can be characterized in two fundamental ways—in terms of a pure capacity limit and as a capacity for meaningful, integrated stimuli, the latter of which may be impossible to pin down precisely (e.g. Cowan, 2000). In a highly influential 1956 paper, Miller reviewed absolute judgment experiments and concluded that working memory has a pure capacity limit of seven plus or minus two *chunks* of information. Miller characterized chunks as related bits of information that become integrated in memory. According to Miller, chunking consists of “organizing bits of information into familiar units... since memory span is a fixed number of chunks; we can increase the number of bits of information that it contains simply by building larger and larger chunks, each one containing more information than before” (p. 91). Miller suggested that humans use this chunking strategy to increase the amount of information that can be held active and processed at one time.

Subsequent research has demonstrated that chunking can be a very powerful memory aid. Chase and Simon’s (1973) study of chess players is one of the earliest and best known demonstrations of the increase in working memory capacity made possible by chunking. They compared the chess game recall of a master, mid-level, and novice chess players’ memory for meaningful and nonmeaningful stimuli—for chess pieces in game-play and unrealistic chessboard configurations, respectively. Participants performed two tasks that Chase and Simon referred to as a perception and a memory task. In each trial of the perception task, the master, mid-level, and novice chess player replicated a chessboard configuration while glancing at the source board as infrequently as possible. In memory task trials, players viewed the chessboard for five seconds and then attempted to replicate from memory the configuration of pieces on the left half of the board. The results indicated that more chess pieces were correctly placed by masters ( $M = 16$  pieces) than by mid-level ( $M = 8$  pieces) or novice ( $M = 4$  pieces) players for mid-game configurations. In the perception task, Chase and Simon identified chunks of recalled chess pieces by noting each glance at the original chessboard and each pause of more than two seconds between the placement of two chess pieces. They found that the expert’s average chunk size for game-play configurations ( $M = 7.7$  pieces per chunk) but not for unrealistic configurations was larger than average chunk size in the low experience groups ( $M = 5.5$  and  $5.7$  pieces per chunk).

More recently, Cowan (2000) has suggested that Miller’s 1956 estimate of pure working memory capacity is inflated. Cowan’s review of working memory capacity research, he claims, supports a pure (i.e., without benefits from rehearsal or existing knowledge structures) storage capacity of three to five chunks. Chen and Cowan (2009) have since refined this estimate to three chunks based on a study in which participants are asked to recall words from lists of different lengths presented under conditions of articulatory suppression (repeating “the”) and following training that focused on gaining familiarity with certain single and paired words. Research by Gobet and his colleagues (Gobet & Clarkson, 2004; Gobet & Simon, 1996) suggests the number may also be lower than Miller’s estimate for information benefitting from existing knowledge structures. Pointing to the results of their computer-based replication of Chase and Simon’s (1973) study, Gobet and Simon (1996) argue that working memory capacity is closer to three or even two chunks. The two researchers used their replication of Chase and Simon’s work to compare two theories: their own *multiple templates theory*, according to which chunks can continue to develop into more complex knowledge structures—templates—that contain slots where chunks can be inserted, and Miller’s *chunking theory*, described above. Their results indicated that the findings of Chase and Simon were affected by the physical limit on the number of chess pieces that could be held in a participant’s hand at one time. This artificial limit may have resulted in an underestimate of chunk size. Gobet and Simon concluded that their findings support template theory; chess players tended to recall two to three interconnected sets, or templates, each containing up to 15 chess pieces. The trend has thus been toward characterizing working memory capacity as small in terms of the number of items—that is, of chunks or templates—and quite large in terms of the amount of information held within each item, particularly in experts.

A related body of theory and research suggests that working memory capacity may get an additional boost from long term memory. The concept of a *long-term working memory* (LTWM) was first introduced by Ericsson and Kintsch (1995) to explain evidence suggestive of an expanded capacity in working memory that allows people, especially experts, to process and work with a great deal more information than should be possible. They suggest that this unexpectedly large amount of actively held information may reflect the use of long-term memory to keep relevant information readily available to working memory. Ericsson and Kintsch argue that with sufficient experience in a given domain, knowledge structures that support performance in that domain form and develop in long term memory. The knowledge structures support rapid information retrieval; retrieval so rapid that information seems to be held continuously in working memory when it is actually shuttling back and forth between working and

long term memory. Information that is active in working memory will trigger a retrieval cue and retrieval structure, activating the interconnected knowledge such that it is readily available for use in working memory. A review of expert and expertise acquisition neuroimaging research by Guida, Gobet, Tardieu, and Nicolas (2012) produced support for both LTWM and template theories of memory and the role they attribute to existing long term memory knowledge structures in extending working memory capacity. In that review, Guida et al. report finding consistent evidence across studies of brain region activation patterns in novice learning and expert performance that were indicative of experts' use of long term memory in working memory tasks.

Sohn and Doane (2004) studied the role of long-term working memory in complex cognitive task performance. Specifically, they looked at instructor versus student pilots' performance on a working memory task and a long term working memory task that used cockpit snapshots as stimuli, and correlated the data with performance on a situation awareness task. They found that experts' long-term working memory scores, but not their working memory scores, were higher than novices'. In addition, performance on the situation awareness task was more correlated with long term working memory than with working memory task performance. The results support the notion that a domain-specific development of long-term working memory can overcome differences in working memory capacity.

## Research Goals

The present research was designed to explore experts' memory. In particular, our goals were to gain insight into how much detail about a given information-rich situation experts are capable of recalling and what strategies and organizational patterns support that recall capability. The domain of expertise chosen for this study is air traffic control, an information-intensive work domain with high memory demands. Air traffic control experts participated in both a typical high demand radar scenario and an equally demanding nonradar scenario in which the controller relies solely on radio-transmitted information to monitor and manage traffic (a rare situation but one for which controllers prepare). Recall data were assessed with the goals of obtaining and examining:

- evidence relevant to the use of long-term working memory to supplement working memory capacity,
- evidence relevant to the use of storage structures such as chunks and templates,
- data characterizing the nature of storage structures that may have been used, and
- evidence suggestive of the use of different memory strategies or organizational patterns when information is visual and auditory (*radar condition*) versus only auditory (*nonradar condition*).

Three experienced air traffic control instructors participated in this qualitative study. Experts are frequently studied in small numbers. For example, the influential Chase and Simon (1973) study described above featured only a single expert and Gobet and Simon (1996) had three, four, and six participants in each of the three experience groups they compared. The goal of these studies is not to use data to test a hypothesis about two or more experimental conditions, which requires a large sample size. Rather, the goal is to explore data from a small number of specialized participants deeply and comprehensively in order to gain insight into an improved understanding of expertise.

## Method

### Participants

Three male retired air traffic controllers between the ages of 60 and 69 with 23 to 34 years of experience controlling high levels of air traffic in a radar environment volunteered for this experiment. Specific demographic data are not provided to ensure participant anonymity. Each participant took part in four 28-minute air traffic control scenarios.

### Experimental Design

The experiment was a two-level within subjects design. The independent variable was control mode (radar vs nonradar). In each control mode condition, the participant performed two scenarios matched for difficulty. Difficulty was not determined by the number of aircraft but by the complexity of control decisions within a scenario. Nonradar control is considered more difficult than radar control by most controllers, and the simulator scenarios difficulty assessment algorithm takes this into account. Each scenario featured three recall trials, two embedded and one at the end of each scenario, for a total of six recall trials in each control mode condition.

### Facility

The experiment was conducted in the En Route Air Traffic Control Lab at Embry-Riddle Aeronautical University in Daytona Beach, Florida. The Federal Aviation Administration (FAA) has certified these labs as official training labs. Students completing their degrees in Air Traffic Management use these labs daily. The R-side (Radar side) position is composed of two 20-inch computer monitors, a communications panel (on the right side of the desktop of each position; see Figure 1), a QWERTY data entry keyboard resembling keyboards used in FAA Air Route Traffic Control Centers, a trackball mouse, a headset with microphone, a push-to-talk foot pedal, and a handheld push-to-talk key. The monitors are vertically stacked in a custom-built cabinet, with the top monitor containing all meteorological information, flow control information, and



Figure 1. Each en route air traffic control simulation system is composed of two positions, a radar side (R-side) on the left and a data side (D-side) on the right. Participants in this study worked in the R-side position. During the nonradar condition, they were supported by an assistant in the D-side position.

active Military Operations Areas (MOAs), while the bottom monitor consists of the radar scope. Two flight progress strip bays sit to the right of the R-side position.

### Confederates

In each scenario, a pseudo pilot sat at a pseudo pilot workstation on the opposite side of the lab from the participant and behind a divider. The pseudo pilot voiced communications for the pilots of all simulated aircraft in the scenario and controlled the aircraft by inputting all aircraft instructions given to pilots by the expert controllers. Pseudo pilot communication to controller was scripted; it appeared in a prompt box when the pseudo pilot clicked on each aircraft. The pseudo pilot was instructed to give no additional information before, during, or after the testing. The pseudo pilots were student laboratory assistants paid by the Air Traffic Control Department at ERAU who had completed the nonradar class. A single pseudo pilot supported all participants across all four scenarios. The pseudo pilot workstation includes a 20-inch computer monitor, a standard Dell keyboard, and a pair of headset/microphone headphones. A foot pedal and hand-activated control for push-to-talk communication activated communication.

In the Radar condition, a data controller (known as the *D-Side controller*) sat to the right of the participant. His purpose was to sequence flight strips according to arrival time for the Radar controller. This D-side role is standard in everyday air traffic control operations and they did not support the participant in any other way. This confederate controller had the same equipment as the primary controller.

### Procedure

The participants were instructed to direct traffic as they normally would. They were informed that each situation was going to be paused twice and that at each pause and at the scenario's end they were to recall the current traffic situation and the characteristics of each flight quickly. Participants were told that each time the scenario pauses or ends, the display would go blank, a blank airspace map would be placed on the desktop in front of them, and they should immediately begin writing recalled aircraft and aircraft details on the map. The following recall instructions were read to each volunteer: "In no particular order, please recall the current traffic situation including items such as aircraft identification, speed, altitude, direction of flight, hand-offs pending or completed, point-outs, arrivals, departures, and any conflicts". Each recall trial had a time limit of four minutes.

After being familiarized with the experiment's procedure, the participants were given a set of flight progress strips and instructed to begin when ready. Each scenario was paused for a recall trial at two times, each time randomly selected from within the windows of 8–12 and 16–20 minutes into the scenario. Two video cameras, focused on the air traffic control radar display and desktop, were used to record each recall trial.

A method used to assist experts in verbalizing cognitive work is the verbal retrospective think-aloud protocol. At the end of each participant's final scenario, the participant was asked to verbally walk-through (recount) his most recent recall trial and to describe the "how" and "why" of the recall process to the extent able. The response was videotaped.

### Data Analysis

#### Traffic Situation Recall

Videotaped recall sessions were examined and the recalled details were transcribed and compared with the corresponding traffic situation. The number and types of aircraft details recalled and patterns in recall were assessed and compared across the radar and nonradar conditions.

#### Retrospective Think-Aloud Data

Retrospective think-aloud data were transcribed and coded (categorized) to characterize responses as implicating the use of either visuospatial or verbal encoding of information. Every data extract in each transcript was assessed by two independent coders to determine whether it referred to the processing or recall of visuospatial information, verbal information or neither and each was coded accordingly. Through a discussion of conflicting codes, code reconciliation was achieved using Cohen's kappa.



## Results

Tables 1 and 2 show the number of aircraft correctly recalled by condition and expert across all six recall trials. The tables show that more aircraft tended to be recalled in the radar than nonradar condition. Because recall trial times were randomly chosen across the two scenarios of each condition, the number of aircraft in the airspace varied (from 2 to 11) across trials.

Aircraft recall order was assessed across recall trials, control mode conditions, and experts. In many trials, the participant first recalled the aircraft he had communicated with most recently; however, communications transcripts were not available to allow the exploration of the potential pattern. No other patterns in aircraft recall order were noted.

Tables 3 and 4 show the percentages of recalled aircraft for which each of four aircraft characteristics were correctly recalled. Included in these tables are the data tag elements speed and altitude, but no other data tag elements (such as point of origin, destination, aircraft type, or heading) because experts never recalled them.

Aircraft call signs were always reported during recall trials, but all three experts reported obtaining call sign by looking up at the flight progress strip of each aircraft after first recalling it by position. They also reported it to be the hardest to remember of the characteristics recalled. Expert 3 sheds light on this during the think-aloud session when he stated, "You don't have to rely on memory for that. It is either A, on your strip or B, on your strip and on the display, so I usually don't try to retain the call sign." Although the use of the flight progress strip could be viewed as compromising the results of this testing, this researcher views it as informative regarding the relationships between memory contents and environmental cues.

Column 4 in Tables 3 and 4 shows the percent of times an assigned action was recalled. If an aircraft was not performing a controller-assigned action at the time of recall, there was no assigned action to recall; however, the percentages in the assigned action column are for all aircraft, not just those that had an assigned action and so are artificially low. (Communications data were not recorded and so aircraft with assigned actions could not be identified.) Consistent with the frequent recall of assigned action, Expert 1, in his think-aloud session, identified assigned action ("what the aircraft were doing") as the

Table 1  
Average number and percent of aircraft recalled across the six recall trials in the radar condition.

Expert	Mean Recalled (SD)	Percent Recalled
Expert 1	7.33 (1.97)	83%
Expert 2	6.50 (1.87)	91%
Expert 3	5.80 (1.47)	71%

Note. The number of aircraft in the airspace varied across recall trials.

Table 2

Average number and percent of aircraft recalled across the six recall trials in the nonradar condition.

Expert	Mean Recalled (SD)	Percent Recalled
Expert 1	5.33 (2.25)	72%
Expert 2	4.50 (1.52)	69%
Expert 3	3.50 (0.83)	60%

Note. The number of aircraft in the airspace varied across recall trials.

easiest information to recall, noting that it is the crux of the job. Assigned action is also central to understanding and remembering the air traffic situation according to the think-aloud description given by Expert 3: "to remember the traffic situation [I usually] remember line-of-flight. I put myself in the pilot's seat. When I issue a clearance, as I am issuing it I am flying it... To know the aircraft's point in space, speed, altitude, and whether it is climbing or descending, is what is important to me."

### Static vs. Dynamic Information

Aircraft characteristics were independently categorized by three researchers into two categories: static and dynamic. Characteristics were characterized as static if they never changed, such as call sign, origin, destination, and type of aircraft. Dynamic characteristics either changed regularly or reflected the development of an aircraft's status or position. They include characteristics such as altitude, speed, hand-off status, point-out highlighting, and departure/arrival status. As shown in Tables 2 and 3, in both control mode conditions experts only recalled dynamic characteristics.

### Organization of Recall

Figures 2 and 3 show the order in which each of the three experts tended to recall aircraft position, call sign, data tag contents, and assigned action. Data tag information recalled was altitude, speed, or both. For the present analysis, the two attributes were grouped together. The percentages shown between each set of nodes in the two figures represent the frequency with which, for each of the three experts, call sign was recalled after position, the frequency with which data tag elements were recalled after call sign when call sign was recalled after position, and so forth.

Recall order tended to follow two dominant patterns. Table 5 shows the frequencies with which the two dominant recall orders occurred in each control mode condition. The columns labeled with the code *P->CS->DT* show the percentage of recalled aircraft for which information was recalled in the position (P), call sign (CS), and data tag (DT) succession. Columns labeled with the code *P->CS->A* show the percentage of recalled aircraft for which information was recalled in the position, call sign, and action information succession.

Table 3

*Percent of aircraft characteristics recalled across the six recall trials in the radar condition.*

	Location	Data Tag: Altitude	Data Tag: Speed	Assigned Action
Expert 1	100%	100%	27%	59%
Expert 2	100%	100%	0%	48%
Expert 3	88%	78%	15%	58%

As seen in Table 5, data tag information tended to be recalled earlier and more reliably in the Radar condition whereas assigned action was recalled earlier and more reliably in the Non-Radar condition.

#### *Modality Comparison: Recall of Visual-Spatial vs. Verbal Information*

As can be seen by comparing the data in Tables 1 through 5, experts tended to recall more visual information (position and altitude) in the Radar condition and more auditory information (assigned action) in the Nonradar condition. Retrospective think-aloud data were assessed to determine if the two conditions also differed with respect to the format in which information was encoded. Think-aloud data were divided into excerpts that were coded to characterize responses as indicative of the use of a visual-spatial or verbal code. The analysis of inter-coder reliability conducted on the pre-reconciled codes produced a Cohen's kappa coefficient of .65, which is considered a fair to good level of agreement (Banerjee, Capozzoli, McSweeney, & Sinha, 1999). Table 6 presents the percent of data extracts in each condition to suggest each type of encoding.

The results of the categorization of think-aloud data extracts demonstrate that in the radar condition, information tended to be encoded and processed in a visuospatial form. The picture is less clear for the nonradar condition. Notably, our expectation that information would be encoded and processed primarily using a verbal code was not borne out.

## Conclusions

In Gobet and Clarkson's (2004) research, experts' ability to recall recent moves as well as "attack" (action) moves were greater than recall for other moves. This tendency for the recall of relevant information supportive of task goals held true in the current study; the experts were able to recall aircraft information that pertained to the movement, separation, and action of aircraft within the airspace. As

seen in Tables 2 and 3, recall for these elements was high across both control conditions. In general, dynamic aircraft information tended to be recalled, whereas static information did not.

The data patterns for the recall of aircraft and aircraft characteristics shed light on experts' strategies, which they have adapted through years of training and experience. The recall patterns found here are consistent with the notion that templates are used to support the storage and retrieval of potentially large quantities of information. More specifically, all data were recalled on an aircraft-by-aircraft basis; that is, information seemed to be recalled as a series of cohesive streams of information about specific aircraft. This quote from the think-aloud data of Expert 3 is consistent with the recall of an integrated set of information: "I didn't really recreate the airspace; I recreated the traffic situation, the approximate position of the aircraft and, in particular, their line of flight or their projected position." This description, combined with observed recall patterns, suggests that the experts' memory storage structures might take the form of aircraft-centered chunks organized within one large template that represents the active portions of the airspace. This conceptualization of a single airspace template is consistent with Cohen (2000) and Gobet and Simon's (1996) re-characterization of chunks as templates and their conclusion that working memory can hold three to five and two to three templates, respectively, of unlimited size.

The organization of the data is also compatible with Klein, Phillips, Rall, and Peluso's (2006) *Data-Frame Theory of Sensemaking*. In the terminology of that theory, the aircraft-related information elements are assigned to preexisting *slots* in each aircraft-centered *frame*, and each aircraft frame might be used to fill slots in an airspace-based frame. According to this theory, the recall patterns suggest a frame (a representation of the airspace) containing an average of six smaller frames (representing individual aircraft), each of which contains four slots (for position, call sign, data tag information, and assigned action).

Table 4

*Percent of aircraft characteristics recalled across the six recall trials in the nonradar condition.*

Expert	Location	Data Tag: Altitude	Data Tag: Speed	Assigned Action
Expert 1	100%	88%	0%	78%
Expert 2	100%	74%	0%	52%
Expert 3	100%	95%	0%	100%

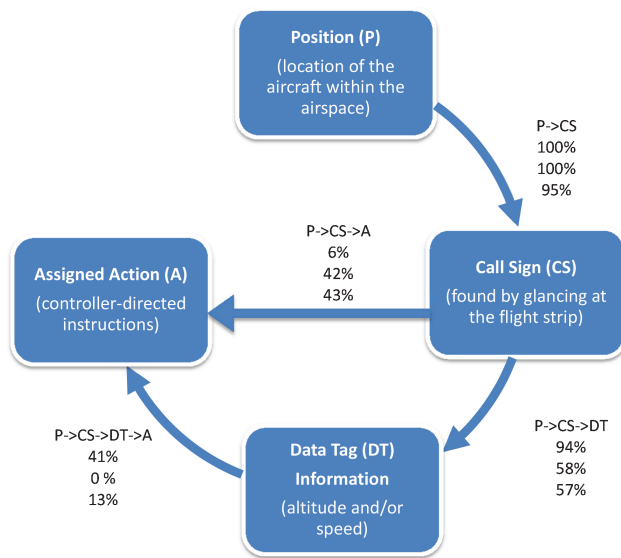


Figure 2. Recall order in the Radar Control Mode condition. Each set of values indicates the percent of aircraft recalled by Experts 1, 2, and 3, respectively, from top to bottom of each stack of values. Moving clockwise from top, the values indicate the percent of aircraft for which position then call sign were recalled; for which position, then call sign, then data tag information, then assigned action were recalled; and for which position, then call sign, then assigned action were recalled.

Observation of the recall process suggested that each recalled aircraft characteristic primed the next item for recall and therefore each recalled item served as a key or

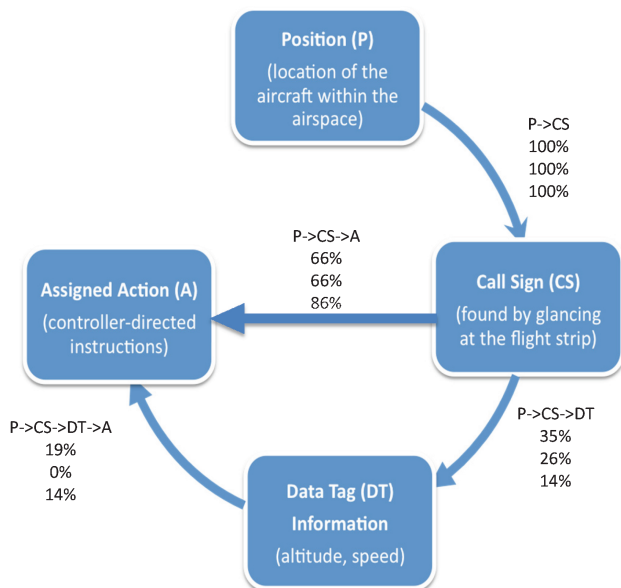


Figure 3. Recall order in the Nonradar Control Mode condition. Each set of values indicates the percent of aircraft recalled by Experts 1, 2, and 3, respectively, from top to bottom of each stack of values. Moving clockwise from top, the values indicate the percent of aircraft for which position then call sign were recalled; for which position, then call sign, then data tag information, then assigned action were recalled; and for which position, then call sign, then assigned action were recalled.

Table 5

The percent of aircraft data recalled in each of the two dominant orders.

Expert	Radar Condition		Nonradar Condition	
	P→CS→DT	P→CS→A	P→CS→DT	P→CS→A
Expert 1	94%	0%	41%	66%
Expert 2	58%	30%	26%	66%
Expert 3	57%	40%	14%	86%

Note. The code P-CS-DT represents the recall order position-call sign-data tag. The code P-CS-A represents the recall order position-call sign-action.

trigger for the next recalled item. The sequential triggering of elements for recall is consistent with Ericsson and Kintsch's (1995) theory of long-term working memory and its proposition that expert performers rely on knowledge in long-term memory and a system of retrieval cues to expand working memory capacity. According to their theory, memory structures, i.e. the templates or frames holding the expert's knowledge of the airspace, are held in long-term memory, not working memory. Their contents are selectively activated and brought into working memory—perhaps one aircraft at a time or even one aircraft characteristic at a time. Data to support this notion of priming are limited to the consistent order in which information types were recalled and observation of participants' body language and patterns in pauses and verbalizations. Observed pauses in recall suggested that experts were actively pulling information into working memory, one set at a time, versus holding multiple sets in working memory from where they might tend to be reported as a single cohesive stream of information.

## Implications

Patterns found in this research may have implications for training strategies useful to controllers in training with the FAA in building better knowledge templates and in starting to build those templates earlier. Educational material could be presented in ways that directly map to or otherwise support the development of meaningful template-based knowledge structures that may aid students in both learning and remembering greater amounts of information.

Additional implications may be found for the Next Generation of air traffic control technologies (NextGen), being developed to, among other things, increase the aircraft tracking capability and capacity of air traffic controllers. This means more aircraft on a display, more tracking, and more decision-making. One issue being considered is reducing the amount of information displayed in the alphanumeric data tag to avoid visual overload. The patterns found in this research suggest that very few information items are actually used to control air traffic. Location, identification, speed, and altitude are the essential pieces of information that participants recalled in the current investigation. Potentially, these are the core pieces

Table 6

Percent of think-aloud protocol data extracts suggesting the use of verbal versus visual-spatial encoding.

Expert	Radar Condition		Non-Radar Condition	
	VisuoSpatial Format	Verbal Format	VisuoSpatial Format	Verbal Format
Expert 1	83% (10)	17% (2)	44% (4)	56% (5)
Expert 2	75% (3)	25% (1)	60% (3)	40% (2)
Expert 3	82% (9)	18% (2)	81% (13)	19% (3)

Note. Only data extracts referring to the verbal or the visuospatial nature of information were extracted from the think-aloud data. Parenthesized numbers represent the total number of categorized extracts in each condition and information format.

of information that should be included in a minimized data tag.

This research also has implications for personnel selection. The Air Traffic – Selection and Training Exam (AT-SAT) is the standardized test that all air traffic control applicants who have received a Tentative Offer (of employment) Letter (TOL) must complete with a score of 70% or greater to continue to be eligible for employment. Applicants that pass this test are divided into three categories: 70.0 to 84.9 is qualified, scores of 85.0 and above are classified as “well qualified.” A third category that has yet to be defined or assigned is “most qualified”. The AT-SAT test consists of eight sections and takes seven hours to complete. One segment features a recall test called “Letter Factory.” In this test, participants are asked to maintain four factory conveyer belts of varying speed. At the end of each belt is a box to be filled. The boxes are different colors as are the letters displayed on the conveyer belts. The purpose is to fill each box with the letters A, B, C, and D of the corresponding color. Other letters are presented on the belts, but the participant must click a quality control button off to the side of the screen when these unwanted letters are presented. Each conveyer belt has a quality control line that indicates the point at which the letters must be placed before they are discarded and unusable. Additionally, at random times throughout the hour-long session, the screen will go blank and a question about the current situation will appear. Questions ask which conveyer belt had the most letters on it, which box was most full, and which belt was moving the fastest.

The results from the present investigation, combined with the findings of Sohn and Doane (2003) discussed earlier, suggest that the letter factory test should be adapted to include information that is relevant to the work domain; in this case, air traffic control. Presenting scenarios such as those used in the current research and pausing each scenario to raise scenario-based questions would increase the sensitivity of the test for assessing the potential to become an effective air traffic controller, which is influenced by pre-existing knowledge, i.e., knowledge in long-term working memory.

In conclusion, the data patterns produced by this work suggest implications for the future of air traffic control

and our understanding of expertise and expert memory capacity. The light shed on experts’ recall capabilities and patterns has implications for the training of novice controllers, radar displays, NextGen technologies, and suggestions for the AT-SAT examination of all potential air traffic controllers. The results also shed light on the knowledge organization strategies experts use and the amount of information they can actively maintain.

## References

- Banerjee, M., Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics*, 27(1), 3–23. <http://dx.doi.org/10.2307/3315487>
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81. [http://dx.doi.org/10.1016/0010-0285\(73\)90004-2](http://dx.doi.org/10.1016/0010-0285(73)90004-2)
- Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology*, 62(7), 1420–1429. <http://dx.doi.org/10.1080/17470210802453977>
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–185. <http://dx.doi.org/10.1017/S0140525X01003922>
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245. <http://dx.doi.org/10.1037/0033-295X.102.2.211>
- Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four, or is it two? *Memory*, 12(6), 732–747. <http://dx.doi.org/10.1080/09658210344000530>
- Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, 31, 1–40. <http://dx.doi.org/10.1006/cogp.1996.0011>
- Guida, A., Gobet, F., Tardieu, H., & Nicolas, S. (2012). How chunks, long term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: A two-stage framework. *Brain and Cognition*, 79(3), 221–244. <http://dx.doi.org/10.1016/j.bandc.2012.01.010>
- Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2006). A data/frame theory of sense making. Proceedings of the 6th International Conference on Naturalistic Decision Making.
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 101(2), 343–352. <http://dx.doi.org/10.1037/h0043158>
- Sohn, Y. W., & Doane, S. M. (2003). Roles of working memory capacity and long-term working memory skill in complex task performance. *Memory & Cognition*, 31(3), 458–466. <http://dx.doi.org/10.3758/BF03194403>