Reduction of Cognitive Workload through Information Chunking

Michael J. Kahn, Kay C. Tan, and Robert J. Beaton
Displays and Controls Laboratory
Department of Industrial Engineering and Operations Research
Virginia Polytechnic Institute and State University
Blacksburg, Virginia USA

Two experiments were conducted to determine whether grouping of icons on complex graphic displays reduces information processing loads, as measured by the Subjective Workload Assessment Technique and errror rates. In Experiment 1, between 2 and 25 symbols were presented on a computer display. Participants were asked to chunk symbols under class labels and store these labels in short-term memory. Two different display formatting variables were tested: spatial proximity grouping of icons was manipulated across three levels, while temporal grouping was manipulated across two levels. Results suggest that display grouping helps operators organize, encode, and store information into task relevant chunks and, in turn, reduces subjective workload and error rates. Experiment 2 was similar to Experiment 1, except that participants were required to remember individual icon names (i.e., participants were asked to remember as many as 25 item names). Results suggest that for chunk formation, storage, and parsing tasks, display grouping may reduce subjective workload, but not error rates.

INTRODUCTION

Electronic displays often provide large volumes of symbolic information relating the status of dynamic system and environmental variables. Systems which employ these complex information displays include air-traffic control, process control, CAD/CAM, and military Horizontal Situation Displays. When using such displays, operators often must integrate, interpret, and act upon information rapidly and accurately. However, because of the volume of information and the rate at which it must be processed, it is possible for operators to become inundated and overloaded. Thus, the operator can be a limiting factor in overall system performance (Goodstein, 1981; Kuperman and Wilson, 1986).

In his now-classic paper, Miller (1956) noted that most individuals can hold 7 ± 2 chunks of information in short-term memory (STM), and that this span of memory is independent of the amount of information (i.e., number of bits) held in each chunk. Miller added, "Since the 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 chunk containing more information than before" (Miller, 1956, p. 93). In the current report, chunking is defined as the cognitive process of organizing information into multi-bit clusters appropriate to solving a particular task.

Chase and Simon (1973) found that expert chess players encoded the location of game pieces into STM more quickly than did novice players. Apparently, expert players are able to group several chess pieces into cognitive clusters or chunks of information. For example, an expert might use the chunk label "castle formation" to encode and store the locations of a king, and three pawns. Conversely, a novice player might encode the same information as four distinct or separate chunks of information.

Conceptually, the process of chunking might proceed as follows: An individual notices a relation among a number of items and assigns a short-hand label (e.g., "castle formation") to that relation. The short-hand labels are stored in STM while the component items (e.g., type and location of individual chess pieces) are held in long-term memory (LTM). To recall the items, the individual first recalls the chunk label from STM, then uses this label to prompt recall of the component items (Chase and Simon, 1973).

Charness (1985) suggested that an expert, in any field, is an individual who maintains a library of learned information chunks (i.e., patterns) and appropriate responses to these chunks. Expertise may provide at least three benefits. First, the expert perceives incoming information in terms of known chunks, hence information can be organized and understood more quickly. Second, the expert knows the appropriate response to most chunks, hence is not forced to calculate new responses "on the fly". Implementation of pre-learned responses generally reduces response times and error rates. Third, by storing labels in STM and using these labels to prompt recall of items held in LTM, the expert is able to increase the amount of situation specific information which can be recalled.

The current research examined the effects of various display formats on non-expert participants' ability to form and utilize information chunks during task completion. It was hypothesized that displays which spatially and/or temporally group icons into task-appropriate clusters would provide the benefits associated with subject matter expertise. Specifically, it was hypothesized that temporal and/or spatial facilitative grouping would aid chunking and, in turn, enhance performance during a time stressed class label extraction task (Experiment 1) and an item memorization and recall task (Experiment 2).

EXPERIMENT 1

Method

Ten male undergraduates participated in this experiment for pay. A Sternberg-type (Sternberg, 1966) method was used for presenting icons and probing STM. During a trial, 2 to 25 iconic symbols were presented briefly on a computer display. The displayed items were sampled from a larger icon set (64 total icons) representing eight unique icon classes (e.g., "military", "kitchen items", "animals", etc.). For example, the "animal" class included the icons: *fish*, *frog*, *turtle*, *snake*, *bird*, *dog*, *bear*, and *elephant*. While viewing the icon presentation during a trial, the participant was asked to remember each icon class name. For example, if the icons: *tank*, *radar*, *missile*, *fork*, *breadboard*, and *cup* were displayed, the participant was required to remember the class names "military" and "kitchen items". Following the icon presentation, the screen was blanked and a probe icon was presented. The participant pressed either a "yes" or "no" button to indicate whether the probe icon was a member of a class which had been presented during that trial. In the example given above, if the probe icon was *spoon*, the participant should have responded "yes" to indicate that *spoon* belonged to the icon class "kitchen items".

Participants viewed each of 144 factorial conditions during a session, and each participant completed one practice session followed by three data collection sessions during the experiment. The 144 trial conditions were determined by a five-factor, within-subjects design. The five factors were: Probe Type (2 levels: positive or negative), Number Of Icon Classes (4 levels: 2, 3, 4, or 5 classes), Number Of Icons Per Class (3 levels: 1, 3, or 5 icons displayed per class), Spatial Proximity (3 levels: random, medium, or tight spacing of displayed class members), and Presentation Order (2 levels: parallel or Chunks-In-Sequence [CIS]). The factorial condition presented for a trial was selected randomly, without-replacement.

Taken together, Number Of Icon Classes and Number Of Icons Per Class provided the the total number of displayed icons (i.e., the display complexity). For example, if a trial called for two Icon Classes with three Icons Per Class, then the icons *tank*, *radar*, *missile*, *fork*, *breadboard*, and *cup* might have been displayed.

The Spatial Proximity and Presentation Order factors controlled the grouping of icons on the display. Spatial Proximity described the spacing of icons within the same class on the display screen. At the "random" (non-chunked) level, icons were assigned randomly across the entire display surface. At the "medium" level, class center points were assigned randomly to locations on the display screen, then each icon was assigned randomly within a 4.5 cm radius of its respective class center point. At the "tight" level, each icon was assigned randomly within 1.5 cm of its class center point.

The Presentation Order factor described the temporal grouping of displayed icons. At the "parallel" level, all icon classes were displayed simultaneously during the trial. At the "CIS" level, the icon classes were presented sequentially, but all members within a class were presented simultaneously. That is, icons belonging to one class would be presented and subsequently blanked, then icons belonging to a second class would be presented and subsequently blanked, etc. Presentation times were balanced across the parallel and CIS conditions such that participants were allocated 800 msec to view each icon class. For example, if icons representing two classes were presented in parallel, these icons would be displayed for 1600 msec. If icons representing two classes were presented in CIS, icons belonging to one class would be presented for 800 msec and blanked; then the icons belonging to the second class would be presented for 800 msec and blanked.

The dependent measures, response accuracy, response time (between probe onset and participant response), and SWAT rating (Reid, Potter, and Bressler, 1989), were collected for each trial. SWAT ratings range from 0 to 100, with 100 representing the highest subjective workload.

Results and Discussion

Each dependent measure was subjected to a separate five-factor Analysis of Variance (ANOVA) procedure. Response time results do not alter interpretations based upon SWAT ratings and error rates, and will be reported in a future article.

Display grouping. In general, display grouping reduced both subjective workload and error rates for the label extraction task.

As Spatial Proximity increased from random to medium to tight, participants were able to extract class labels with less subjective workload, $F(2,18) = 26.61$, $p < 0.0001$. Similarly, SWAT ratings are lower for CIS than for Parallel Presentation Order, $F(1,9) = 26.13$, $p < 0.0006$. Finally, as shown in Figure 1, the two-factor interaction between Spatial Proximity and Presentation Order is significant, $F(2,18) = 25.72$, $p < 0.0001$.

A Newman-Keuls test, indicates that Spatial Proximity does not affect SWAT ratings under CIS Presentation Order, $p > 0.05$, but that SWAT ratings decrease as Spatial Proximity increases under Parallel Presentation Order, $p < 0.05$. The Random-Proximity/Parallel-Presentation-Order condition may produce high workload since participants often sampled icons representing already-sampled classes (i.e., had already stored that class label in STM). Redundant within class sampling costs cognitive effort while providing no additional information. In the Tight Proximity condition, same class icons were close together and often separated from other icons. These grouping cues provided "boundaries" which participants could recognize and utilize easily. Redundant within class sampling was reduced, thus reducing subjective workload.

CIS Presentation Order also provided boundaries between classes; however, these boundaries were temporal rather than spatial. It may be that, for CIS presentations, SWAT ratings were equal across levels of Spatial Proximity because temporal grouping was fully sufficient. That is, if classes were separated temporally, the physical spacing among icons did not affect subjective difficulty. Finally, in Figure 1, note that the Tight-Proximity/Parallel-Order condition produces low SWAT ratings similar to CIS Presentation Order conditions. It appears that, for the label extraction task, subjective workload was equal whether boundaries between classes were spatial or temporal.

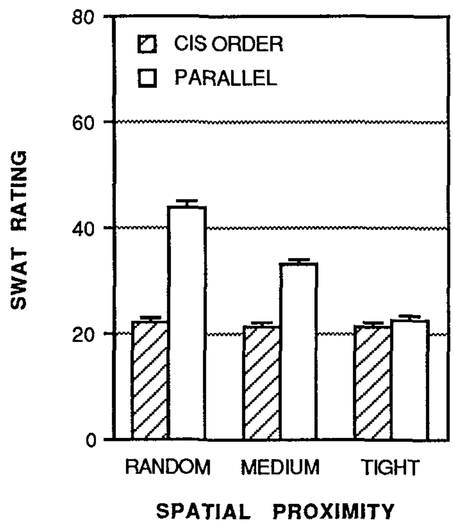

Figure 1. Two-factor interaction of Spatial Proximity by Presentation Order, collapsed over subjects (N = 10), Sessions, Number Of Icon Classes, Number Of Icons Per Class, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars indicate + 1 standard error of the mean ($S_{\bar{X}}$).

In a manner similar to SWAT ratings, error rates decrease monotonically as Spatial Proximity increases, $F(1,9) = 5.19$, $p < 0.0166$. The two-factor interaction between Spatial Proximity and Presentation Order also is significant, $F(2,18) = 25.72$, $p < 0.0001$. A Newman-Keuls test indicates that more errors occur in the Random-Proximity/Parallel-Order than in any other condition, $p < 0.05$. Again, it appears that the Random-Proximity/Parallel-Order condition provided no boundaries so that participants expended time sampling icons which belonged to already sampled classes, thus often failing to sample other classes.

Display complexity, interaction between display complexity and display grouping. In general, increases in display complexity produce increases in subjective workload and error rates for the label extraction task. In addition, as display complexity increases, benefits of facilitative grouping become more pronounced.

SWAT ratings increase monotonically as Number Of Icon Classes increases, $F(3,27) = 41.60$, $p < 0.0001$, and Number Of Icons Per Class increases, $F(2,18) = 30.51$, $p < 0.0001$. Error rates also increase monotonically with increasing Number Of Icon Classes, $F(3,27) = 10.87$, $p < 0.0001$, and increasing Number Of Icons Per Class, $F(2,18) = 3.27$, $p < 0.0615$. There are numerous SWAT and error rate two- and three-factor interactions between/among the Number Of Icons Per Class and variables controlling display grouping (Spatial Proximity and Presentation Order). An example is the SWAT rating, two-factor interaction between Number Of Icon Classes and Spatial Proximity, $F(4,36) = 26.17$, $p < 0.0001$. Figure 2 shows that Spatial Proximity has no effect on SWAT ratings when the Number Of Icons Per Class equals one, $p > 0.05$, or when Spatial Proximity is Tight, $p > 0.05$. However, as the Number Of Icons Per Class increases, workload costs due to low Proximity also increase, $p < 0.05$ (across relevant comparisons).

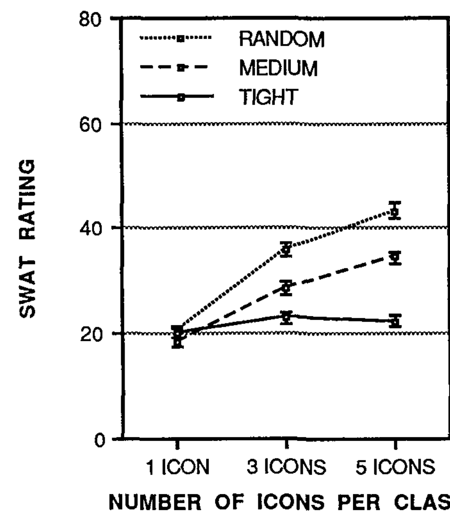

Figure 2. Two-factor interaction of Number Of Icons Per Class by Spatial Proximity, collapsed over subjects (N = 10), Sessions, Number Of Icon Classes, Presentation Order, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars indicate $\pm$ 1 $S_{\bar{X}}$.

Figure 2 data can be explained as follows: first, when the Number Of Icons Per Class equals one, Spatial Proximity has no effect. Spatial Proximity describes the spatial relationship of icons belonging to the same class. Hence, when there is only one item per class, it is not meaningful to manipulate spatial distances between items belonging to that class (i.e., displays are equivalent regardless of the level of Spatial Proximity). Second, when Spatial Proximity was tight, facilitative spatial grouping provided easily discernable boundaries between icon classes, so participants easily extracted the class name regardless of the Number Of Icons Per Class. However, as Proximity decreased participants had either more ambiguous or non-existent spatial boundaries to help chunk icons belonging to the same class. In this case, increases in the Number Of Icons Per Class caused increases in task difficulty.

EXPERIMENT 2

Method

The methodology used in Experiment 2 was similar to that used in Experiment 1, except that ten new participants were required to remember each icon presented during a trial rather than just the icon class names (i.e., participants were asked to remember as many as 25 item names). Since Experiment 2 often required participants to recall numbers of displayed items in excess of the 7 $\pm$ 2 item STM capacity, suggested by Miller (1956), it was believed that proficient performance would require participants to organize information into chunks, store and recall the chunks, and then parse those chunks into component items. It was hypothesized that display formats which grouped same class icons would facilitate efficient chunking, storage, and parsing as necessary for the item memorization and recall task.

Results and Discussion

As in Experiment 1, each dependent measure was subjected to a separate five-factor ANOVA procedure.

Display grouping. Overall, display grouping reduces subjective workload, but not error rates, for the item memorization and recall task.

As in Experiment 1, as Spatial Proximity increased, participants completed the task with less subjective workload, $F(2,18) = 29.41, p < 0.0001$. Similarly, SWAT ratings are lower for CIS than for Parallel Presentation Order, $F(1,9) = 20.86, p < 0.0014$. Finally, the two-factor interaction between Spatial Proximity and Presentation Order is significant, $F(2,18) = 22.84, p < 0.0001$. A graph of these results would differ from Figure 1 in overall SWAT levels (SWAT ratings were higher for Experiment 2), but not in the data pattern.

Unlike Experiment 1, error rates do not differ significantly between CIS and Parallel Presentation Orders, $F(1,9) = 0.41, p = 0.5380$, nor is there a trend suggesting a reduction for CIS Presentation. Similarly, error rates do not vary significantly among levels of Spatial Proximity, $F(2,18) = 0.43, p = 0.6599$, nor is there a trend suggesting an error rate reduction for Tight Proximity.

It is interesting that treatments which group icons produce significant and relatively large reductions in subjective workload, while producing no reduction in error rate. One possible explanation is that facilitative display grouping helped participants organize information into chunks, thus reducing subjective workload, yet participants were unable to effectively parse these chunks into component items; thus, error rates remained constant. This reasoning is consistent with Experiment 1 results which indicate that facilitative grouping formats produce lower subjective workload and error rates for a chunk-creation task similar to that required as part of the Experiment 2 task.

Since previous research has indicated that chunking reduces error rates, while the current research does not, it is possible that there are two kinds of chunks: STM-LTM chunks and STM chunks. STM-LTM chunking, was described earlier in this article (e.g., Chase and Simon, 1973; Miller, 1966), and describes the use of a smaller number of shorthand labels stored in STM to access a larger number of items stored in LTM. As discussed earlier, STM-LTM chunking may be available only to experts who maintain a library of learned information patterns and appropriate responses to these patterns. The second type of chunking, STM chunking, describes a situation where individuals organize information in STM, but do not have well-learned patterns. Therefore, information can not be "off-loaded" to LTM.

For example, if Experiment 2 participants had used STM-LTM chunking, each STM label would always have accessed a specific subset of icons. Thus, "mammal" might always access *bear*, *dog*, and *elephant*, while never being used to access more, fewer, or different icons. Upon seeing the *bear*, *dog*, and *elephant* icons, the participant would immediately encode the STM-LTM chunk "mammal"; hence, one STM label would be used to store three icons. In this case CIS Presentation Order or Tight Proximity might have helped participants encode and store the icons *bear, dog, fork, plate, elephant, spoon,* as two chunks of information with shorthand labels: "mammal" and "place setting". Instead, it is likely that facilitative grouping helped Experiment 2 participants organize icons into STM chunks. For example, CIS Presentation Order or Tight Proximity might have helped a participant organize the icons listed above into the STM string *bear, dog, elephant, fork, plate, spoon.* This STM string is organized so that similar icons are contiguous in STM, yet knowing that "animals" were presented would not allow a participant to recall whether *bear*, *dog*, and *elephant* or *bear*, *dog*, and *turtle* were presented. Therefore, this STM chunking does not allow information off-loading to LTM.

Display complexity, interaction between display complexity and display grouping. Overall, display complexity increases subjective workload and error rates for the item memorization and recall task. Also, as display complexity increases, subjective workload, but not error rate, benefits of facilitative grouping become more pronounced.

SWAT ratings increase monotonically as the Number Of Icon Classes increases, $F(3,27) = 118.26, p < 0.0001$, and Number Of Icons Per Class increases, $F(2,18) = 99.83, p < 0.0001$. Error rates also increase monotonically with increasing Number Of Icon Classes, $F(3,27) = 15.58, p < 0.0001$, and Number Of Icons Per Class, $F(2,18) = 76.09, p < 0.0001$. There are numerous SWAT two- and three-factor interactions between/among display complexity variables (Number Of Icon Classes and Number Of Icons Per Class) and display grouping variables (Spatial Proximity and Presentation Order). There are no error rate results indicating that display grouping effects vary with display complexity. This latter result is not surprising as no main effects indicate that facilitative grouping reduces error rates for the Experiment 2 task.

The SWAT two-factor interaction between Number Of Icons Per Class and Spatial Proximity is shown in Figure 3. A Newman-Keuls test indicates that SWAT ratings are equal when the Number Of Icons Per Class equals one, $p > 0.05$, and increases with increases in Number Of Icons Per Class, $p < 0.05$, and decreased Spatial Proximity, $p < 0.05$ (across most relevant comparisons). These results are consistent with those of Experiment 1.

However, unlike Experiment 1, subjective workload for Tight Proximity increases with Number Of Icons Per Class. One explanation is that, in Experiment 1, same-class icons provided redundant information, and facilitative grouping allowed participants to easily eliminate sources of redundant information, while in Experiment 2, same-class information was related, but not redundant. Hence, facilitative grouping reduced, but did not eliminate, subjective workload associated with increasing numbers of same-class items.

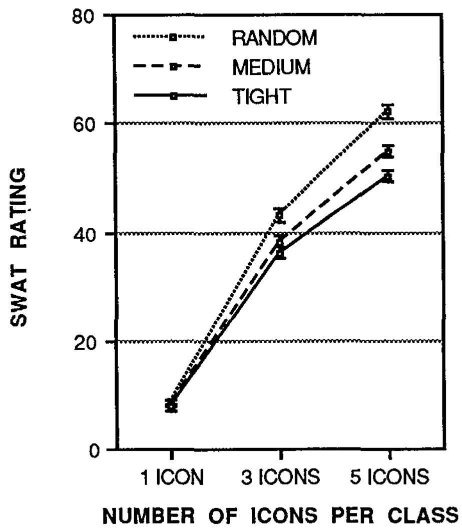

Figure 3. Two-factor interaction of Number Of Icons Per Class by Spatial Proximity, collapsed over subjects (N = 10), Sessions, Number Of Icon Classes, Presentation Order, and Type, on SWAT ratings in Experiment 1 ($p < 0.0001$). Error bars indicate $\pm$ 1 $S_{\bar{X}}$.

CONCLUSIONS AND RESEARCH NEEDS

Results indicate that spatial and temporal grouping can reduce cognitive workload on high information density tasks. For a task requiring class label extraction (Experiment 1), spatial and temporal grouping reduced both subjective workload and error rates. For a task requiring recall of large numbers of displayed items (Experiment 2), spatial and temporal grouping reduced subjective workload, but not error rates (i.e., recall of component items). These results are highly encouraging, especially as it is likely that chunk recognition is more important to operator performance than component item recall. Also, it is possible that for participants who are expert in a task, spatial or temporal grouping might reduce error rates; that is, display grouping would help experts recognize chunks. The experts would then call upon known rules to store and parse these chunks. Conversely, future research might reveal that experts are able to perceive information chunks so easily that display grouping would provide no subjective or performance benefit.

Finally, future research should address two potential concerns which might limit the applicability of CIS grouping. First, while CIS grouping might allow operators enhanced understanding of individual chunks, it may inhibit understanding of relations across chunks. Second, in CIS grouping, information is presented at a computer paced rate. Operators may at times feel information is presented too slowly or quickly, or fail to see information which is displayed and then disappears.

REFERENCES

Chase, W. G., and Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Charness, N. (1985). Aging and problem solving performance. In N. Charness (Ed.), *Aging and human Performance* (pp. 225-229). Chichester: Wiley

Goodstein, L. P. (1981). Discriminative display support for process operators. In J. Rasmussen and W. B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 433-449). New York: Plenum Press.

Kuperman, G. G., and Wilson, D. L. (1986). *A human-centered technology for advanced bomber crew stations* (Report No. AAMRL-TR-86-052). Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Reid, G. B., Potter, S. S., and Bressler, J. R. (1989, July). *Subjective Workload Assessment Technique (SWAT): A User's Guide (U).* (Report No. AAMRL-TR-89-023). Wright-Patterson AFB, OH: Armstrong Aerospace Medical research Laboratory.

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153*, 652-654.

ACKNOWLEDGEMENTS

This research was supported by the Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio, under contract number F33615-88-D-0532 (Task #2), to Dr. Robert J. Beaton.