Design and Evaluation of a Cognitively Engineered Systems Monitoring Display

Michael J. Findler
Wright Ste Research

Dr. S. Narayanan
Wright State University

Jeff S. Collier

Dr. William Marshak

DESIGN AND EVALUATION OF A COGNITIVELY ENGINEERED SYSTEMS MONITORING DISPLAY

Michael J. Findler
Wright State Research Institute
Dayton, OH USA
Dr. S. Narayanan,
Wright State University
Dayton, OH USA
Jeff S. Collier
Sytronics, Inc
Beavercreek, OH USA
Dr. William Marshak
Sytronics, Inc
Beavercreek, OH USA

A new format was derived from a Visual Thinking cognitive psychology paradigm and permits easy understanding of multiple system parameters with different directions and unit scaling. This new "Sprocket" format allows rapid cross check, characterizing multiple failure thresholds, and easy detection of out-of-tolerance conditions and a gestalt state awareness. The format was evaluated in a dual task, aviation-oriented experiment.

The advent of the all glass cockpit in aircraft makes new display formats possible beyond the traditional discrete gauge display that dominated the first 100 years of aviation. Graphics allow greater flexibility in information display, but is not without its dangers. A British B737 accident was attributed to an aircrew unfamiliar with a new engine display format shutting down the wrong engine in response to an engine fire. New format designs must be cognitively compatible with both the aircrew's mental model of aircraft function, at least consistent with existing displays affecting reading through transfer of training, and provide significant advantages in reading accuracy, speed, and reduced display space.

Visual Thinking

The Visual Thinking phrase was first coined by Arnheim (1969) to describe the relationship of perception and cognition. Inherent in vision is the ability to preprocess data and recognize visual patterns.

> Vision is not perception and perception is not thinking. The mind gathers information and processes it. Note that I said information, information is data plus meaning. Before the mind conveys the information your eyes must observe it, and some preprocessing needs to be done to turn this data into information. (Arnheim 2004)

The key to Arnheim's thesis is that vision and thinking are not necessarily disjoint concepts. When a person perceives an object with their eyes, before deep thoughts about the object can be conceived, the simple sight of that thing at least causes classification (placing the object in the context of other objects like it). For instance, if you see a cat, before any separate thinking is performed about the cat, it has already been placed in the category of "cat". This is a particularly useful cognitive trait to have when that cat is a dangerous one that needs to be fled from, such as a tiger.

Arnheim contends the idea of visual thinking is an old one, going back to the ancient Greek philosophers: Plato, Socrates and Aristotle. These philosophers were the first to make a distinction between perceiving and reasoning, mainly because perception from direct senses could not always be trusted. (We have all experienced "our eyes playing tricks on us", or heard tales of mirages in the desert.) Reasoning was considered to be the "correction of the senses" and the "establishment of truth".

It can be reasonably argued that Arnheim's Visual Thinking is an almost instantaneous pattern classification. It is not the perception of the object that classifies the object, but rather the very well travelled mental pathways that react with almost lightening quick classification. The perceiving of the object (cat) does not require new neurons to fire off and create new paths; the existing short pathways have always succeeded previously.

Design of the Visual Thinking Sprocket

Physiologically, the eyeball as an information-gathering instrument scans the world under the guidance of cognitive attention centers. The eyeball fixates on a region of interest. An image is buffered and scanned, like a massively parallel computer, to find objects within the image through feature extraction. Once extracted, these objects are serially scanned at about 25 items per second. Since the eye scans quickly, reacquiring a new image about 10 times a second, only four to twelve objects are recognized before the eye jumps to another fixation. These physical boundaries must drive the design of cognitively sensitive displays.

Furthermore, when designing a display, two attributes must be balanced: the **overview** of the situation and the **details** within the situation. The overview is a qualitative "aspect of data preferably acquired rapidly and even better, pre-attentively; that is, without cognitive effort" (Spence 2007). A well designed overview display uses visual cues that are acknowledged to be pattern classifier aids so information "pops out" at the operator. On the other hand, details are quantitative and should only be presented to the operator on an as needed basis, i.e., when the operator requests more in-depth information, presumably because of the overview display observations.

Within the design of the Visual Thinking Sprocket display, primary attention is devoted to the overview pattern classifier aids. A design that presents an overview of a situation must be designed simply and stress those features that can be pre-attentively processed. According to Ware (2004), features that can be pre-attentively processed can be organized in the following categories:

- Form: Line orientation, line length, **line width**, line collinearity, **size**, **curvature**, **special grouping**, blur, added marks, numerosity
- Color: **Hue**, Intensity
- Motion: Flicker, **Direction of Motion**
- Spatial Position: 2D position, Stereoscopic depth, convex/concave shape from shading

The features in **bold** were the pre-attentive cues the Visual Thinking Sprocket attempted to model. Examining ***Figure 1***, one can see the intentional feature implementation on the initial single-threshold Visual Thinking Sprocket design. This Visual Thinking Sprocket was intended to be a decision support aid within a larger flight simulator.

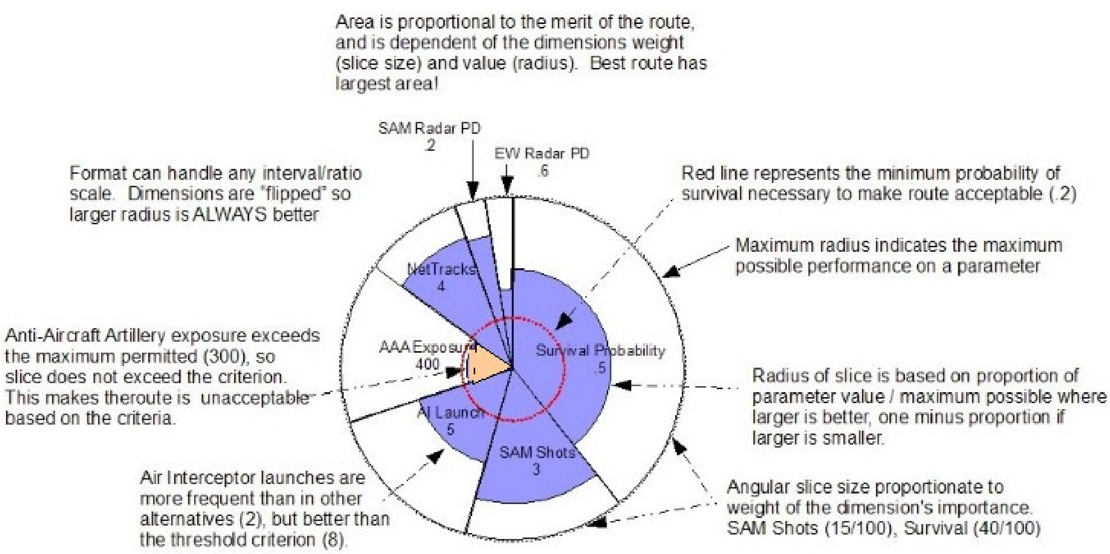


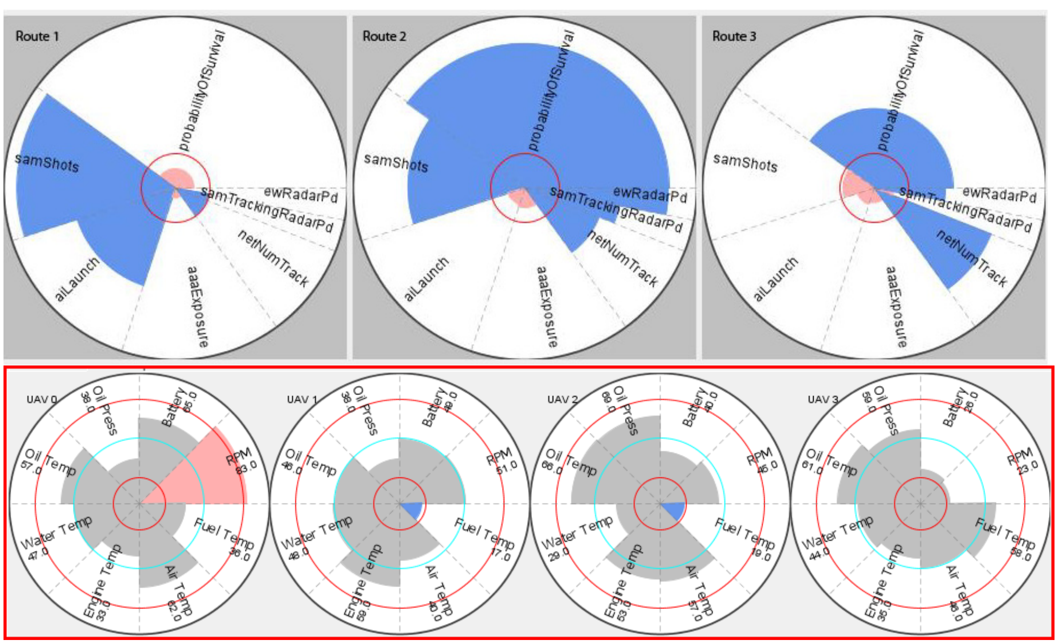
Figure 1. Early drawing of a multi-dimensional, multiple scaled decision support display. The raw detail data display is visible by mouse roll-over of the slice.

Encoded into this initial Visual Thinking Sprocket were (1) angular slices proportional to the weighting of the dimension; (2) acceptability of specific dimensions (pink – unacceptable, blue – acceptable); (3) individual dimension "health" or "preference" (larger colored area is always better); (4) slices nearer the red tolerance line are less optimal, those nearer the maximum radius are deemed near optimal; (5) labels naming individual dimensions and their associated current values; (6) a normalized rescaling of the dimensions; and (7) the global preference of the decision – bigger sprockets are better than smaller sprockets. Finally, if the operator wanted more information about a specific dimension, a simple “mouse-over” displays the detailed raw data behind the image.

From Figure 1, one can see why the resulting circular figure is called a sprocket, with geared teeth of varying length, resembling the tooth embellished wheel that drives a chain, or in this case, cognitive understanding. Experiments were performed to examine the viability of using Visual Thinking as the cornerstone in designing displays. Figure 2 illustrates two versions of a Visual Thinking Sprocket, each of which can be instantiated through the same software library.

Figure 2.A The decision support aid is a static single-threshold two-color sprocket – it compares three alternate routes that an operator may select. The operator has to choose which route alternative was best based on the presented dimensions, called the Figures of Merit (FOM). For this display, bigger is always better, so Route 2 is the best route. This is obvious without placing exact values on the display. The sprockets were instantaneous snapshots of the route alternatives. Finally, dimensional weights are displayed in the slice angular subtend so dimensional contribution to total area is clearly expressed.

Figure 2.B The dynamic double-threshold three-color system monitoring display. The operator is monitoring four unmanned aerial vehicles (UAVs). In this example each of the UAVs has serious health issues, for example UAV_1 and UAV_2 have Fuel Temperatures that are below the minimum threshold, while UAV_0 has an RPM that exceeds the upper threshold. The amount that UAV_1 and UAV_2 fail to reach the minimum threshold is illustrated by the size of the wedge – UAV_2 “barely” fails to reach the minimum threshold, indicated by the numeric value and less white space in between the wedge and red lower threshold ring. UAV_3 is “healthy”, since it is all gray. Best is not the biggest or smallest area; but rather, the most circular gray pie. Optimal is indicated by the light blue ring. Color coding is meant to mimic American water faucets, e.g., blue (cold/low) and red (hot/high).



A

B

Figure 2. Static single-threshold (A) and dynamic double-threshold (B) Visual Thinking Sprocket designs

The new sprocket format is designed to achieve greater situation awareness (SA) as a basis for the decision. Endsley and Kiris (1995) defined three different levels of SA. Level 1 (SA1) deals with the “perception” of elements of the environment. Level 2 (SA2) describes the comprehension of those elements and indicates a deeper level of cognitive comprehension. Level 3 (SA3) refers to the ability to use that comprehension to make predictions based on them. When operators choose between generated routes, they necessarily make an SA3 assessment to predict which route will best accomplish the mission. The Visual Thinking design paradigm encourages greater comprehension because it reveals not only which route is superior, but why it is superior. It facilitates integration of the various dimensions via the rescaling of the dimensions and translation into the area representation.

The Visual Thinking Sprocket was compared to a text table (similar to the current UAV display technology (circa 2007) used in a Predator Ground Control Station (GCS) Variable Information Tables (VIT) and bar charts (considered chosen as a first naive attempt at a "graphical user interface"). It is understood that these are discrete interfaces, not integral interfaces, which is because there exist no integral interfaces for multiple UAVs.

Experimental Design

The experiment employed the Sprocket as a decision aid to choose which of three alternative missions were best based on weighted criteria. The criteria included dimensions like probability of survival, number of missile shots, minutes of radar exposure, etc.; measures differed in their direction and scales. This trio of routes was presented in a two factor repeated-measures design with full-model partitioning. Each image was generated from the same data, i.e., a bar chart, text table and sprocket image were generated from datasets 1, 2, …, 12. Each subject was shown a series of generated images of the Figures of Merit (FOM) for three alternative paths' and a question. The subject responded to the first question (rank order the three routes from best to worst), and then the second question was displayed, and so on – the presented image did not change and remained visible during the questioning (SPAM – Situation Present Assessment Method). The four questions and levels of situation awareness are listed in Table 1. The questions were designed to explore specific decisions considered typical within a multi-UAV mission.

Table 1. Questions asked for each set of generated images.

	SA Level	Question	Possible Answers
Q1	SA2	Rank order the routes [Best to worst]	1-2-3, 1-3-2 2-1-3, 2-3-1 3-1-2, 3-2-1
Q2	SA1	Do any of the routes meet the all minimum criteria?	Yes, No
Q3	SA3	Which route is Best if Dimension X is dropped? [Where X was chosen from among the 4 top weighted available dimensions]	1, 2, 3
Q4	SA1 & SA2	Which route has Best Dimension Y? [Where Y was chosen from all available dimensions]	1, 2, 3

Results and Conclusions

The accuracy and response time for each question must be examined. To choose this new display, the results of the experiment must show that it is better than the alternatives (current text based display or naive graphic display) in accuracy and/or response time – preferably both. Furthermore, the subjects should find the new display "intuitive".

Looking at the "correctness of answer" per question data first, the generated data was examined prior to presentation to the subjects to determine the "correct answer" to each question with a weighted combination of the parameters. The operator's "correctness" response was then defined as whether the operator responded with the pre-calculated correct answer. A correct response was assigned a value of 1 and a wrong answer was assigned a value of 0. The sum of the correct answers was then used as a measure to determine the cognitive ease of use for each display type – each question had a maximum score of 12 points for correctness.

For example, the Minimum Criteria question (Q2) asks whether all of the mission alternatives presented to the operator meet the minimum criteria on all dimensions: (1) the possible answers are `Yes` or `No`, and (2) the correct answer is `Yes` (the three mission statements are all valid).[1] Then for each operator that answered `Yes`, a

[1] The Text Table and the Sprocket color coded a failure of each dimension, i.e., if the "Probability of Success" dimension failed to reach the minimum value, then the text (pie piece) would be colored red. For the Bar Chart, no color coding was attempted.

counter would be incremented by 1. The maximum value the counter could reach is 12, so if the final counter value was 12 (out of 12), then 100% of the subjects responded with the correct answer. If ½ of the subjects responded correctly, then the final value would be 6 (out of 12).

Table 2 summarizes the statistical results for each question. Significant statistical differences in accuracy were found in questions 1, 2 and 3.

Table 2. Analysis of accuracy results

Question	Statistic	Tukey *post hoc* analysis	Conclusion
Q1 Rank Order	[F(2, 46)=11.70, p<0.0001]	Tukey Grouping / Mean / N / Block A / 7.17 / 24 / Sprocket B / 5.38 / 24 / Bar Chart B / 4.96 / 24 / Text Table	Sprocket significantly better
Q2 Meets Criteria	[F(2, 46)= 7.82, p=0.0012]	Tukey Grouping / Mean / N / Display A / 11.29 / 24 / Sprocket A / 10.96 / 24 / Text Table B / 9.92 / 24 / Bar Chart	Bar chart signicantly worse
Q3 Drop One Dimension	[F(2, 46)= 14.03, p<0.0001]	Tukey Grouping / Mean / N / Display A / 9.17 / 24 / Sprocket B / 7.75 / 24 / Bar Chart B / 7.17 / 24 / Text Table	Sprocket significantly better
Q4 Best Dimension	[F(2, 46)= 2.03, p<0.1432]		No significant difference

Now that the statistical results of the accuracy among the questions have been examined, we looked at the response times for each question. If the main effects of Display were found to be significant, there is a statistical difference in the response times among the displays, a Tukey *post hoc* test was performed to find the significance. As anticipated, the main effect of Trial (image presentation order) was found to be significant. Images presented at the beginning of the trial took significantly longer to interpret than those presented at the end of the trial, irrespective of display type, also known as a learning curve. But the most interesting finding is that the display type significantly affected the learning curve slope (Display×Trial). The statistical results are summarized in Table 3.

Table 3. Analysis of response times

Question	Source	Statistic (**Bold** is significant)	Conclusion (from Tukey *post hoc* [not shown])
Q1 Rank Order	Display	**[F(2, 46)= 30.91, *p*<0.0001]**	**Sprocket** significantly better
	Trial	**[*F*(11,253)=17.44, *p*<0.0001]**	Learning curve existed
	Display×Trial	**[*F*(22,506)=7.99, *p*<0.0001]**	Display type significantly affected learning curve
Q2 Meets Criteria	Display	**[F(2, 46)= 29.86, *p*<0.0001]**	**Sprocket** significantly better
	Trial	**[*F*(11,253)=13.27, *p*<0.0001]**	Learning curve existed
	Display×Trial	**[*F*(22,506)=2.08, *p*<0.0001]**	Display type significantly affected learning curve
Q3 Drop One Dimension	Display	[F(2, 46)= 2.96, *p*<0.0626]	**No significant difference**
	Trial	**[*F*(11,253)=9.37, *p*<0.0001]**	Learning curve existed
	Display×Trial	**[*F*(22,506)=1.98, *p*<0.0001]**	Display type significantly affected learning curve
Q4 Best Dimension	Display	**[F(2, 46)= 8.85, *p*<0.0001]**	**Text table** significantly better
	Trial	**[*F*(11,253)=12.41, *p*<0.0001]**	Learning curve existed
	Display×Trial	**[*F*(22,506)=8.04, *p*<0.0001]**	Display type significantly affected learning curve

Because of significance of the Display×Trial interactions, the time series results are graphed in ***Figure 3***. The graphs show the average time each question took to be answered by the operators. Note that the images were presented in a Latin Squares random order. The x-axis labels are the order the images were presented, i.e., first image presented, second image presented, etc. The experiment was considered long enough at approximately 15-20 minutes per display for the learning curves to be examined. Without fail, the initial reaction to the bar charts across all displays had the longest interpretation times and the sprocket tended to have the lowest mean response time over the course of the experiment. However, interpreting which sprocket had the best value for a given dimension (Q4) had the text and sprocket supplying similar response times.

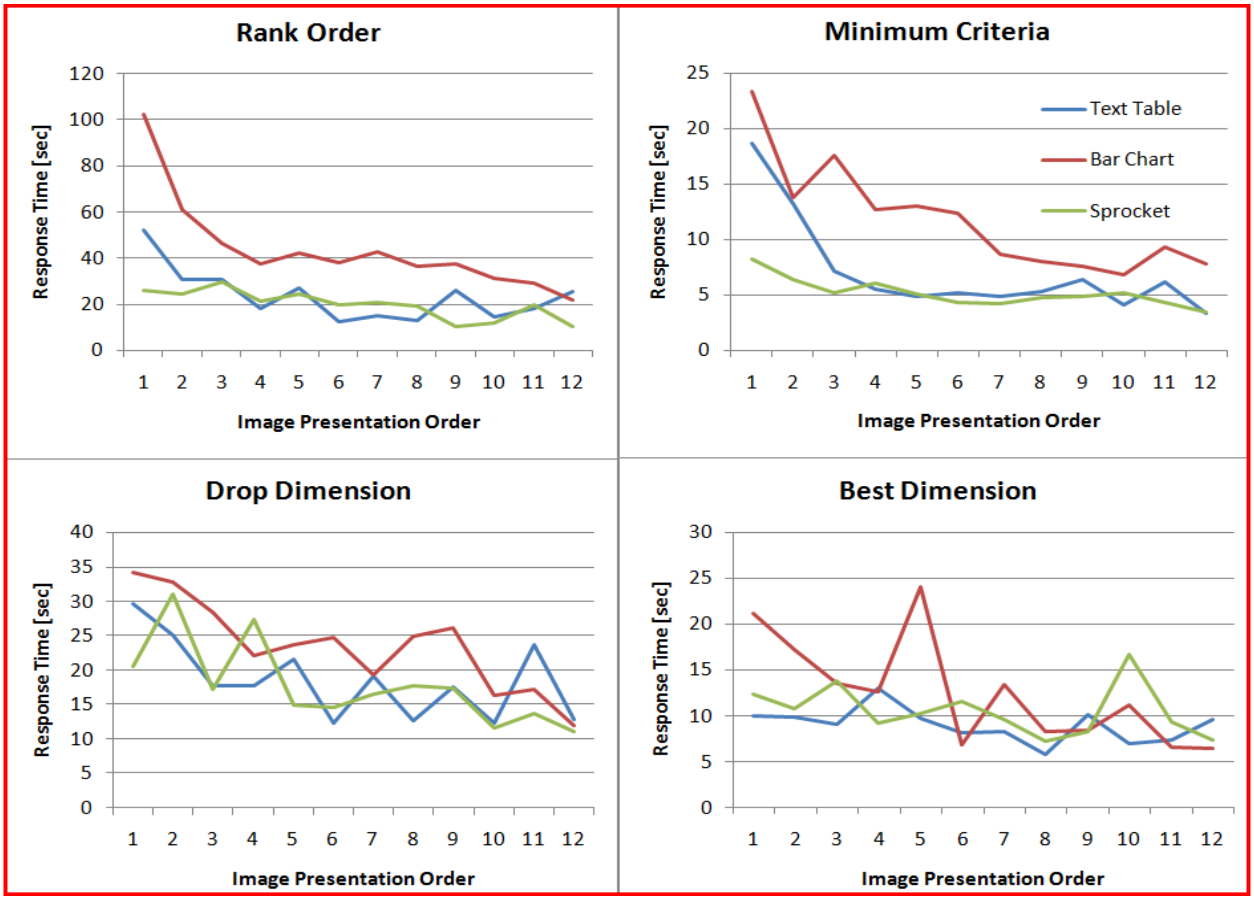


Figure 3. Learning curve for each question

The Sprocket display format offers a spatially compact, concise, single instrument that displays the overall quality of each route and allows ready comparison of their merits. Rescaling the direction and range of dimensions masks their exact value details (which can still be displayed by our software as a "roll-over"), converting it into a less precise value in the angle and radius in the Sprocket. The format exploits the cognitive-perceptual ability to compare size or areas of objects into the cognitive understanding of a mission route's absolute and relative merits. Even more exciting is the Sprocket and the Visual Thinking paradigm from which it was developed, represents only one member of a class of new display formats that exploits the connection between perception and cognition.

References

Arnheim, R. (1969). Visual Thinking, University of California Press.

Arnheim, R. (2004). Visual Thinking, University of California Press.

Endsley, M. R. and E. O. Kiris (1995). "The Out-of-the-Loop Performance Problem and Level of Control in Automation." Human Factors **37**(2): 381-394.

Spence, R. (2007). Information Visualization: Design for Interaction. Essex, England, Pearson Education Limited.

Ware, C. (2004). Information Visualization: Perception for Design. San Francisco, CA, Elsevier.