JENNIFER M. HOCKO

Merging Verbal and Visual Media in Information Graphics

This paper discusses the cognitive processing that users of information graphics must perform in order to understand what is being presented. Because the ways viewers process verbal and visual media differ, so do the tasks for which each medium is best suited. This discussion therefore highlights the strengths and weaknesses of both the verbal and visual mediums, and describes some conditions under which they independently excel. Next, this paper describes how visualizations that combine both visual and verbal media can assist viewers by reducing their cognitive load and taking advantage of viewers' advanced visual processing capabilities. This paper closes by describing a case study where an information graphic is used to support a specific user's task by combining both visual and verbal elements.

Cognitive Processing of Verbal and Visual Elements

There are two aspects to cognitively processing information graphics, whether they contain verbal or visual elements: encoding the information, and storing that information in memory for subsequent retrieval. Humans encode verbal information (that is, spoken language or text) as a series of discrete propositions. These propositions or "symbolic representation[s]...[are] used to construct a mental model" (Rusbult) of the information that can then be stored. Because verbal stimuli prompt an ordered, two-step response (to first translate the stimuli into propositions, then fit those propositions into a mental model), it is not surprising that "it can take a few seconds to hear or read a short sentence" (Ware, pp. 316). Once verbal stimuli is encoded in this manner, the information is stored in an area of memory allocated specifically for verbal codes (Schreiber, et. al., pp 115). If one assumes there is a similar decoding phase associated with verbal information, subsequent retrieval of information from the verbal memory store should require at least the same amount of time.

Unlike verbal information, humans encode visual information in a parallel manner (Ware, pp. 316). Verbal and visual information is encoded differently mainly because of an image's unique ability to simultaneously communicate two types of information, categorized as "feature information" and "structural information." Verdi, et.al. describe feature information as "the visual properties of the individual features," or the icons, labels, colors, sizes, shapes, and so on that are “used to depict features.” Structural information is essentially spatial information, or information that communicates the “distance between features” (pp. 305; Schreiber, et. al., pp. 115). Ware describes spatial information simply as “the way [objects] are laid out in a particular environment” (pp. 312). Because they contain both types of information, images are encoded as “intact units that take up less cognitive space,” and allow “viewers to shift attention across the image space without using all the resources of working memory” (Verdi, et. al., pp. 305). In other words, visual information “allow[s] a more direct construction of a mental model” because it does not require intermediate steps like the construction of discrete propositions (Rusbult). Therefore, even when looking at a visual display that contains many independent details, viewers are able to “comprehend [that] complex visual structure in a fraction of a second, based on a single glance” (Ware, pp. 316). Another consequence of processing visual displays in parallel is that information “[can be scanned] rapidly” and used to “quickly...discover patterns of elements...that are meaningful” (Rusbult). Once visual stimuli is processed in this manner, the information is stored in an area of memory allocated specifically for visual codes (Schreiber, et. al., pp 115). Because they are encoded and stored as “whole units,” the visual information also requires less resources to subsequently retrieve, because both types of information contained within the visual display are “simultaneously available” (Schreiber, et. al., pp. 115).

Contrasting the Use of Visual and Verbal Displays

This section contrasts the use of visual and verbal displays by explaining the picture superiority effect and describing the strengths of visual displays. It also explains that some types of communication are still best supported through the verbal medium.

Explaining the Picture Superiority Effect

Numerous studies have shown that when used in isolation, visual elements are superior to verbal elements for presenting information. This is especially true when viewers are subsequently asked to recall that information. There are two main theories for explaining this “picture superiority effect:” Paivio’s dual-coding theory (DCT), and what is known as the “distinctiveness model” (Kulhavy, et. al., pp. 1-2; Mintzer and Snodgras, pp. 1-2). The DCT proposes that pictures are recalled more frequently and accurately than words because of the two *separate but interconnected* areas in memory that are used for coding and storing verbal and visual information. When only a visual stimulus is presented, a viewer naturally attempts to name the image, thereby coding the stimulus both ways and activating both the visual and verbal stores (David, pp. 184). Visual stimuli are thus said to be "dual-coded." Alternatively, when only a verbal stimulus is presented, the viewer is not typically inclined to do the reverse (that is, form a mental image of that word) and therefore, only codes the stimulus verbally and only activates the verbal store. As a result, "pictures are remembered better than words because pictures are more likely than words to evoke the verbal and image codes...and the storage of two memory traces increases the probability of retrieving a stored event" (Mintzer and Snodgras, pp. 1). The distinctiveness model is a less popular but equally plausible theory that attributes the superiority of pictures to their "highly distinctive visual [semantic] features that allow them to be *uniquely encoded* in memory." In other words, the distinctiveness model advocates that there is some attribute of the visual elements themselves that causes them to be coded differently (and separately) from their verbal equivalents (Mintzer and Snodgras, pp. 1). Although this model agrees that visual and verbal stores are interconnected, it advances the idea that this separate storage (rather than the dual-coding) is responsible for "reducing interference from competing responses...making a correct retrieval more likely" (Kulhavy, et.al., pp. 1).

Besides increasing the likelihood that information will be accurately and quickly remembered, "Cognitive Processing of Verbal and Visual Elements" on page 1 touched on a key strength of visual displays: their ability to effectively communicate "spatial structure, location, and detail" as well as the relationships among collections of concrete objects (Ware, pp. 319-320). The low cognitive load and memorability of visual displays is also a direct result of how images work to "[exploit a human's] physiological strengths...such as pattern recognition and the ability to recognize geometric shapes" (Rusbult). A verbal display used to communicate detailed feature and structural information might take up pages, and still would not provide the same level of understanding because the serial nature of verbal processing places too high a load on working memory.

Possible Explanations for Word Superiority

Because humans' processing of visual displays is far superior to our processing of verbal displays, authors like Ware admit that it can be "difficult to pin down the advantages of words" (pp. 332) but clearly, there are some. First, spoken and written language provides designers of information graphics with a system of symbols that have shared, widely-known meanings, which may make verbal symbols superior to those chosen for visual displays (Rusbult; Ware, pp. 322, 332). Simply stated, the "quality of...communication depends on the...symbols shared by [a designer] and viewer" (Rusbult). Lifetime training in and cultural focus on verbal communication may also mean that verbal displays do not present viewers with as many obstacles to mutual understanding (Ware, pp. 332) because they have become the preferred “cognitive style” for processing information (Sadler-Smith). These factors, therefore, may make viewers more adept at processing verbal information than it would initially appear.

David illustrates in his experiments how the “concreteness [of a concept] has a strong influence on [its] imagistic potential” (pp. 184, 196), indicating that abstract concepts may still be best communicated verbally. An example of an abstract concept is “freedom,” whereas a concrete concept would be one like “hammer” (David, pp. 183-184). Ware also supports the idea that “words are better for representing...abstract verbal concepts” (pp. 319-320). Under certain conditions, it may also be important for viewers of an information graphic to follow a specific procedure; in some situations, failure to do so may have catastrophic results. Thus, the sequential processing that takes place in the mind when information is presented verbally may be perfectly matched to such tasks (Ware, pp. 319-320). Encouraging viewers to consciously process information by using a medium that produces a slightly higher cognitive load may actually help ensure that the proper level of attention is given to a critical procedure. Additionally, a verbal display’s ability to communicate “qualifying information” or specific “conditions” for behavior is unmatched by any visual technique (Ware, pp. 319-320). For all these reasons, “words should [still] provide the general framework for the narrative...and...[be] used for the detailed structure” (Ware, pp. 322).

An Argument for Combining Verbal and Visual Media

Tufte states that “words, graphics, and tables are different mechanisms with but a single purpose—the presentation of information,” and his disapproval of the separation between verbal and visual media is clear when he says “words and pictures belong together” (pp. 180-181). Even if a picture is not redundant to the text it accompanies, studies have shown that the mere inclusion of that picture “produces significant gain in recall” (David, pp. 191; Ware, pp. 324). Others have shown that “instructional procedures...[are] understood better if blocks of text [are] integrated with the diagram...[and] steps read immediately adjacent to relevant visual information” (Schreiber, et.al., pp. 115-116; Ware, pp. 324). Although Tufte is an expert in the field of information graphics and experiment results are solid, it is more interesting to examine *why* words and pictures work together rather than to simply believe it.

Combining verbal and visual media together in a single information graphic is, for the most part, more effective than using either medium alone because the ways in which each "contribute to the construction of a mental model" are "complimentary" (Rusbult). Storing redundant verbal and visual codes in separate stores and activating the link between those stores (as described in "Explaining the Picture Superiority Effect" on page 2) serves to reinforce and strengthen understanding of that information. And in some cases, the disadvantages of one medium are compensated for by including the other. For example, Ware states that the use of "static images by themselves are not effective in providing complex, nonspatial instructions" (pp. 320). Using words and images together to produce a complete and doubly-strong encoding can only be achieved with the proper integration (linking) of the verbal and visual elements of the display (Ware, pp. 319, 333). Such integration prevents viewers from having to perform that task in their working memory, thereby reducing the cognitive load that a mixed display produces (Kalyuga, et. al., pp. 3). Ware's advice for properly linking verbal and visual elements include: (1) "associating words with the appropriate images" (pp. 323); (2) drawing attention to the appropriate visual, based on what the viewer is verbally processing (pp. 326-327); and (3) adhering to Gestalt principles such as proximity to clearly indicate groupings of objects (pp. 323; also Schreiber, et.al., pp. 116).

Another reason for using information graphics with combined verbal and visual elements has to do with cognitive style and learning preferences. Sadler-Smith explains that a person with a "verbalizer" cognitive style "will learn better from textual information whereas imagers will learn better from pictorial information. In each case the processing load is minimized since the mode of presentation is congruent with the individual's habitual mode." Although audience analysis can help designers understand many things about viewers of information displays, it is doubtful that information at this level of individual detail can typically be obtained. Thus, by including both verbal and visual elements in information graphics, an "individual learner's...preferences with respect to abilities, prior knowledge, strategies and interests" can be accommodated (Rusbult).

However, Some Words of Caution

When designers combine verbal and visual elements in an information graphic, they must also be sensitive to how viewers will allocate their attention to them. Schreiber, et. al. found that the use of "considerate text" (defined as text that discusses the features of a graphic "on the basis of the predicted scanning pattern") increases learning and retention of information (pp. 117). The tendency for these scanning patterns (typically "top-down" or "bottom-to-top") may depend on many factors, including cultural reading habits, the dimensionality of the graphic, and so on (Schreiber, et. al., pp.

116-117). Tufte warns that the objective of an information graphic is a serious consideration as well; if the verbal portion of the stimuli is "too effective," it may "tell viewers how to allocate their attention to the various parts of the data display" instead of allowing viewers to freely interpret its content (pp. 182).

There are also cases where using redundant visual and verbal stimuli in an information graphic may not be advantageous. One factor that may produce such a case is the target viewer's level of expertise with the subject matter being communicated. Kalyuga, et. al. found that while the cognitive load for novice viewers is reduced by merging verbal and visual cues (for all the reasons previously described), expert viewers can find this integration and redundancy distracting. In the case of expert users, the use of both verbal and visual elements covering the same information actually increase cognitive load (pp. 2, 14-15) because there is "competition for the limited executive resources needed to control the two types of processing" (Rusbult). Cognitive loads may be increased even further as expert viewers of verbal/visual displays attempt to focus their attention by ignoring one medium or the other. In addition to creating a higher mental workload, viewers may also miss critical information by focusing their attention away from "the most effective processing mode" (Rusbult). Thus, Kalyuga, et. al. provide us with yet another reason why designers of information graphics should know their audience (pp. 16).

Case Study: Providing Directions

This case study focuses on providing the author's mother with directions to the Best Western Terrace Inn in Brighton, MA, for when she drives with her significant other from Dallas, Pennsylvania to visit.

Audience

Debbie and Jim have driven from Dallas, Pennsylvania to visit the author's former home in Connecticut many times before, but they have not yet driven around the new neighborhood near Boston. Jim will be driving while Debbie navigates using the information graphic, so Debbie is the primary audience for the display. Debbie had not traveled much until her 40s, but has become increasingly capable of reading typical road maps and following written directions as she broadens her horizons. (The fact that Debbie is not driving herself is also likely to increase her confidence and comfort in using the directions.) Like her daughter, Debbie is not always good at getting her bearings in unfamiliar places, and has a tendency to lose her focus at times. Debbie is not likely to have experience using the kind of information graphic (map) that is attached.

Objectives and Use Case

The primary objectives of an information graphic that provides Debbie with directions are: (1) to get Debbie and Jim to their final destination and (2) to prevent them from taking incorrect routes. Secondary objectives for the information graphic involve both the audience and the situation in which the graphic will be used—the graphic must also (3) depict only relevant details that allow Debbie to easily recognize her present environment and (4) allow Debbie to quickly locate their location on the graphic after having looked elsewhere.

Rationale for Attached Information Graphic

Naturally, this designer has attempted to leverage the benefits of both visual and verbal media and the ways each is cognitively processed to support the audience, use case, and objectives described. The attached information graphic:

- Provides procedural information as verbal elements in plain, simple English.
- Provides spatial (structural) information as visual elements oriented in a way that is always consistent with how the user will perceive her environment. (In other words, the map need not be turned so that it still makes sense.)
- Provides visual elements with minimal complexity because the context in which they will be studied provides little time for doing so. This is in line with Ware's advice that "simple line drawings may be most effective for quick exposures" (pp. 319-320).
- Describes through visual elements only the relevant details of the journey (except in cases where error-preventing landmarks are used) thereby simplifying the procedure. As Rusbult states, "drawings are intended to be accurate in some respects but not others...[the] amounts and types of simplification [depend] on the...objectives."
- Contains visual and verbal elements such as labels for landmarks (that is, feature information) in areas where there is a high likelihood that users may make an error.

- Arranges each step in the procedure as a separate pane, taking even more advantage of spatial relationships (proximity, top-down and left-to-right processing, and so on) to aid the user in keeping track of where they are in the overall process.
- Facilitates recognition tasks by providing redundant verbal and visual cues within each pane to reinforce the information presented and to reduce the cognitive load placed on the target user.

References

David, P. (1998). News Concreteness and Visual-Verbal Association: Do News Pictures Narrow the Recall Gap Between Concrete and Abstract News? In Human Communication Research, 25(2), pp. 180-201. International Communication Association.

Kalyuga, S., Chandler, P., and Sweller, J. (1998). Levels of Expertise and Instructional Design. In Human Factors, 40(1), pp. 1-17. Human Factors and Ergonomics Society.

Kulhavy, R.W., Stock, W.A., Woodard, K.A., Haygood, R.C. (1993). Comparing Elaboration and Dual Coding Theories: the Case of Maps and Text. In American Journal of Psychology, 106(4), pp. 1-10. University of Illinois Press.

Mintzer, M.Z. and Snodgras, J.G. (1999). The Picture Superiority Effect: Support for the Distinctiveness Model. In American Journal of Psychology, 112(1), pp. 1-22). University of Illinois Press.

Rusbult, C. (1995). Visual Thinking and Visual-Verbal Communication Page. http://www.sit.wisc.edu/~crusbult/methods/visual.htm. Accessed 3 April 2002.

Schreiber, J. B., Verdi, M.P., Patock-Peckham, J., Johnson, J.T., Kealy, W.A. (2002). Differing Map Construction and Text Organization and Their Effects on Retention. In the Journal of Experimental Education, 70(2), pp. 114-130.

Sadler-Smith, E. (2000). The Implications of Cognitive Style for Management Education and Development: Some Evidence from the UK Page. http://www.elsinnet.org.uk/abstracts/aom/sad-aom.htm. Accessed 3 April 2002.

Tufte, E. (1983). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.

Verdi, M.P., Johnson, J. T., Stock, W.A., Kulhavy, R.W., Whitman-Ahern, P. (1997). Organized Spatial Displays and Texts: Effects of Presentation Order and Display Type on Learning Outcomes. In the Journal of Experimental Education 65(4), pp. 303-317.

Ware, C. (2000). Information Visualization: Perception for Design. San Francisco, CA: Morgan Kaufmann Publishers.