VISIBLE LANGUAGE FOR THE EXPRESSION OF SCIENTIFIC CONCEPTS

MIKE ZENDER

KEITH A. CRUTCHER

ABSTRACT

The accelerating rate of data generation and resulting publications are taxing the ability of scientific investigators to stay current with the emerging literature. This problem, acute in science, is not uncommon in other areas. New approaches to managing this explosion of information are needed. While it is only possible to read one paper or abstract at a time, it is possible to grasp concepts presented visually in milliseconds. This suggests the possibility of developing a visual language to represent concepts from a multitude of published papers in an accurate display that is highly condensed, yet readable in seconds.

This paper describes the initial exploration of a visual language approach to the display of concepts found in published scientific papers: in this case, some hypotheses surrounding the etiology of Alzheimer's Disease.

The approach is based on deriving propositions from papers or abstracts, breaking propositions into concept objects, designing a visual object system (consisting of icons, signs, glyphs and combinations) to represent all the objects in the relevant concept space, displaying the objects as a networked constellation and linking the visual display back to the papers from which they came.

University of Cincinnati
Visible Language 41.1
Zender and Crutcher, 23-49

Rhode Island School of Design
Providence, Rhode Island 02903

The ultimate goal is to develop visual language techniques capable of revealing patterns, pathways and conceptual connections not readily apparent from text-based list of findings and using such visual language to make interactive displays that accurately represent large quantities of data in a condensed conceptual form. Such an approach has potential application to any field of study that has a controlled vocabulary.

FACETS OF A PROBLEM

You can see a lot. Thanks to the computer and the Internet, there is a lot more to see, read and comprehend than ever before. According to Lyman and Varian, in the year 2000 the world produced between one and two exabytes (a billion gigabytes) of unique information, about 250 megabytes for every man woman and child on earth (Lyman and Varian, 2000). By 2003 this was five exabytes annually.

Those leading the pursuit of specialized knowledge, whether in science or other areas, are also on the forefront of dealing with this growth of information in dramatic ways. In science, the primary source of new information is peer-reviewed papers and abstracts published in specialized journals. In the area of medical research, the National Library of Medicine manages an online resource known as PubMed that currently hosts over 12,000,000 journal articles. Users can type a text query and retrieve all of the relevant references (with abstracts) based on key words. A recent (10/24/06) PubMed query using "Alzheimer*" as the search term, for example, returned 54,430 citations. The rate at which the literature in this field has increased during the lifetime of one of us (KAC) is shown in the accompanying chart *(figure 1).*

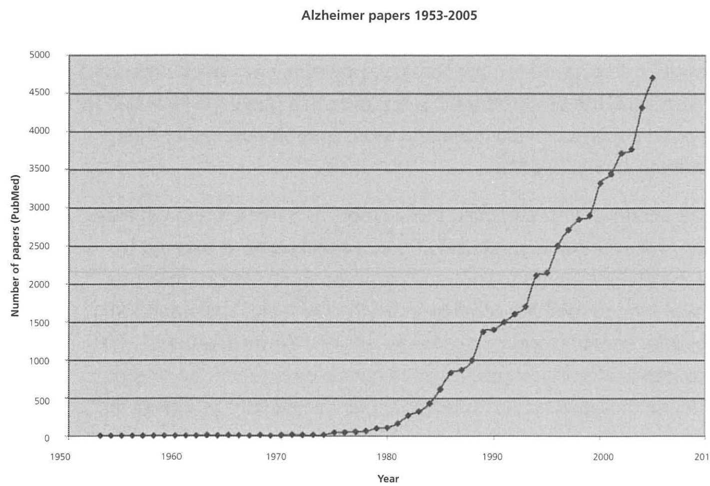

Figure 1.

Let's imagine that a new investigator wants to become familiar with this literature. At fifteen minutes per paper, it would take seven years of reading to get through all 54,000 – allowing for eight hours of reading per day with time off for weekends, holidays and vacations! To master the literature in the area of Alzheimer's Disease, however, would also require reading new papers, which are currently published at a rate close to 5,000 per year. By the time the investigator had finished reading what is available today, the new stack of unread papers would be 35,000 (assuming no further increase in the number of papers per year). Yet staying abreast of what is happening in their field is exactly what scientists are expected to do. Clearly, reading every paper is not viable. This constant growth in information, acute in medicine and science, also occurs in other fields.

As if this explosion of data was not problem enough, biological systems have inherent complexity and relevant data come from various fields of study and various levels of analysis, ranging from atoms to populations. Add to this the desire to compare results from different laboratories and approaches and the need for improved techniques for experiencing information is clear.

The growth in information and the problems this raises have not gone unnoticed. The field of bioinformatics has evolved, in part, to respond to this issue. Various approaches to managing the literature and scientific data have been put forward. However, even when important information has been retrieved in response to a query the information is often displayed in quantities and forms that obscure rather than facilitate understanding.

One approach to overcoming these problems is information visualization. The fact that a great deal of the cerebral cortex is devoted to processing visual information has led scientists to convert data into visual form to facilitate cognition. Graphs and images are common tools for scientific communication (as the first figure illustrates). The advantage of using images for communication is clear from history. The use of glyphs and icons in ancient languages such as Egyptian and Mayan, and even earlier uses of pictures in cave art, such as that found at Lascaux, attest to the utility of symbolic and direct representation of images for communication. The major advance provided by invention of alphabets (and, eventually, the printing press), on the other hand, was the rapidity and efficiency with which information could be stored, reproduced and communicated. The time and labor needed to reproduce images in pre-computer history simply could not compete with the advantages of the alphabet. However, the capabilities of modern computers (including hardware and software advances) now make it possible to reproduce images and recover the advantages of visual language with much less labor and time. Thus, the technological advances that have contributed to the problem of too much data also provide possible solutions.

Pre-computer computational approaches demonstrated the benefits of using visual form to gain understanding from data. Jacques Bertin advanced the idea that through interaction with graphics (visual displays of data), one can gain insight and convert data into knowledge

(Bertin, 1983). Bertin's work predated contemporary computer systems and relied primarily on physical manipulation of paper cards containing drawn graphics, but the principles, rather than being diminished by time, are more applicable than ever with the advent of computationally-based interactive media and data processing.

In the past twenty years, Information Visualization has grown into a specialized field with its own journals, conferences, theoretical basis and research foci (generating its own information explosion!). Books by Jacques Bertin, *The Semiology of Graphics* (1967) and *Graphics Information Processing* (1983) and several works by Edward Tufte (1983–97), though relatively recent, are nevertheless considered to be seminal works in the field. In 1986, the National Science Foundation (NSF) launched a new initiative in scientific visualization and the first Institute of Electrical and Electronics Engineers (IEEE) Visualization conference was held in 1990 (Card et. al., 1999). There is great interest in information visualization in the scientific and medical communities as evidenced by the birth of Informatics sections at Universities and grant funding opportunities in the United States at the National Institute of Health (NIH) and National Science Foundation (NSF) calling for research in information visualization. Scientists, computer engineers, and programmers are entering the fields of bioinformatics to move visualization forward. Most of the contemporary computer-based visualization approaches they develop use some combination of simple graphics and text, such as the link node diagram. These approaches have demonstrated effectiveness yet still often fall short of supplying understanding. Using words as labels to communicate concepts in such visual displays has several inherent flaws: it is language-based, takes time to read and the visual form of the word has nothing to do with the concept or idea it represents.

Even though visualization is defined as a form of communication (DeFanti, Brown and McCormick, 1989), visual communication designers have not been deeply involved. One potential impediment is the relative paucity of collaborations between scientists and graphic designers. To address this challenge, one of us (MZ), a designer, and the other (KAC), a biomedical scientist, formed a collaboration to introduce scientific approaches to information designers and to make scientists aware of the capabilities of visual communication.

CONTEXT FOR A SOLUTION

Recognizing that much was being done to visualize data, the authors wondered whether it might be possible to visually represent the key concepts and ideas found in scientific papers in a more immediate way than text-based approaches. The ability of visual form to summarize large data sets is well established (Tufte, 1983; Ware, 2004). The ability of icons to communicate concepts is similarly well documented and a part of everyday life (Arnheim, 1974). The utility of scientific and mathematical visual notation systems is also commonplace, although these systems, like all sign-based systems, require special learning. We wondered if key concepts in fields with controlled vocabularies, such as medicine, might be efficiently communicated with images such as glyphs or icons, and, if so, whether these images might then effectively illustrate the web of conceptual connections spread across hundreds or thousands of journal articles and papers within a specific area of investigation. If such a system were interactive, we suspect that it might lead scientists to insights more quickly than scanning mountains of papers. If such a system also remained linked to individual papers then such a visual display might be an improved means of exploring a literature database such as PubMed.

Several impediments to a visual solution to a language-based system exist. One problem is language itself. Language is notoriously vague. The meaning of words depends on context. N. T. Wright explains this well in his description of word meanings (Wright, 1992):

First, the meaning of a word (following Wittgenstein) I take to be its use in a context, or an implicit context; that is, its use in a sentence or potential sentence. If I say 'book,' the meaning of this is in doubt until I form a sentence: 'I am going to book tickets'; 'The book is on the desk'; 'The criminal was brought to book.' Even where a word is clearly univocal, we can never rule out possible metaphorical meanings, and in many cases we only know the univocal meaning through experience of sentences in which it has become plain.

In other words, meaning in language is not easily defined. The controlled vocabularies noted below confirm the existence of this problem as they attempt to solve it.

Visual language faces the same challenge. Humans use context to assign meaning to the stimulus of visual perception (Arnheim, 1969). Current icon-based communication systems generally rely on their environmental context, such as an airport or highway, to provide context that helps define meaning. Even so, existing icons do not communicate as precisely as written language. In particular, icon systems, while effective at communicating physical objects, rarely communicate processes and actions (Zender, 2006). This is like a spoken language with no verbs. Attempts have been made to develop universal systems of non-verbal visual communication, notably Isotype, but these have been widely regarded as failures (Lupton, 1989). Investigations into means of expanding the scope and effectiveness of icon-based non-verbal communication are being made (Zender, 2006), but such efforts are still embryonic.

Finally, the conceptual spaces that all languages, visual or otherwise, attempt to describe are not always well defined themselves. Ontologies have different structural qualities and may not easily map onto real-world objects. One familiar analogy is the parent-child relationship found in tree structure ontologies. Each term can be categorized based on known relationships. However, not every field of study has a defined conceptual structure, let alone clear analogies. Furthermore, the relationships between conceptual objects are often not understood until more information is obtained. As a result, it is not possible to strictly define a concept space in advance.

Fortunately, in the sciences in general, and medicine in particular, many of the problems noted above have been addressed. In relation to our proposed question, we identified several developments in science that might support a visual language solution. The issue of language vagueness, for example, can be partially addressed with the tools available through the Unified Medical Language System (UMLS). One of these tools, the Semantic Network, is a system of categorization for concepts that are contained in the UMLS Metathesaurus, which is a database of terms relating to health and biomedicine. The Metathesaurus provides a means to identify and resolve synonyms and a variety of specialized vocabularies that apply to defined contexts. For example, 'to dress' from a doctors' perspective may mean to bandage an incision, whereas from a nursing perspective it may mean to put clothes on a patient in preparation for discharge. In addition, PubMed uses a vocabulary known as MeSH (Medical Subject Headings) to classify and categorize the content of papers. This is an open vocabulary (recent new descriptors have been added, for example, as a result of interest in avian influenza), but it seeks to retain control over the indexing terms used for this database. Another related technology is the growth in Natural Language Processing (NLP) software, which is able to parse electronic texts and correctly identify key words, such as UMLS terms.

In addition to developments to overcome language vagueness, new techniques for visualizing information have evolved. Some, such as Fish Eye Views (Furnace, 1989) clarify hierarchy in large bodies of text while maintaining essential context *(figures 2-5).* Others have focused specifically on visualizing documents. TitleBars uses topics provided by users to find documents and build a visual display in the form of a bar for each document with relevant portions of the bar highlighted in various gray values. "Themescapes is a 'thematic terrain' that communicates the primary themes of a collection of documents and relative prevalence of those themes." (Wise et al., 1995 quoted from Spence, 2001) The terrain metaphor visualizes conceptual structure with higher levels, mountains, being more general conceptually and lower levels, valleys, being more specific. Another approach, self-organizing maps, sometimes called Kohonen maps, represent content as a grid of diverse puzzle shapes whose size and location, in the case of document visualization, represent quantity and conceptual relationship, respectively. All of these approaches combine a graphic and typographic representation with words defining concepts. Telemakus is a more recent system for extracting concepts from journal articles which displays them in a graphic/typographic form (Revere, 2003).

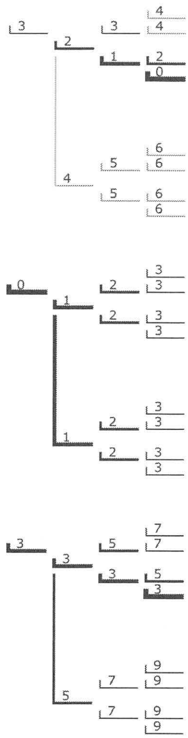

Figure 2. The Fisheye view for tree structured content. The top diagram illustrates the focal point: a data node of interest, given a value of 0. Each step further removed from the focal point is given a higher number: 2 - 6. Furnas calls this the distance from focus, D(.,x), and is calculated thus: D(.,x) = d(.,x). The middle diagram illustrates what Furnas calls the level of detail (LOD), the intrinsic importance of each level of content measured by the distance from the most important root of the tree, being 0. The formula is: LOD(x) = -d(r, x). The bottom diagram illustrates what Furnas calls the degree of interest: the sum of the focal point and the level of detail. In this technique, lower numbers are nodes that are closer to the interest and the key organizational structure of the content.

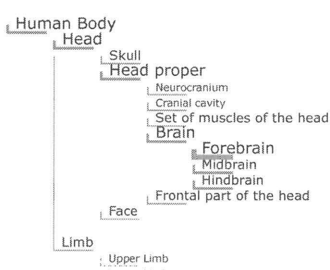

Figure 3. The Fisheye view is applied to a content tree. Assuming the search was on the term "Forebrain" the type sizes and line thicknesses are progressively smaller based on the distance from the search term plus the degree of interest in proportion to the Fisheye formula.

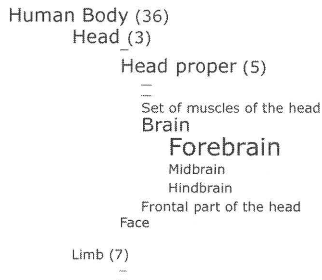

Figure 4. Using the Fisheye technique, content further from the search term (focus) may be eliminated because relevant context is retained by the formula. In this example, numbers indicate how many items are under each tree section and small rules represent content that is not shown.

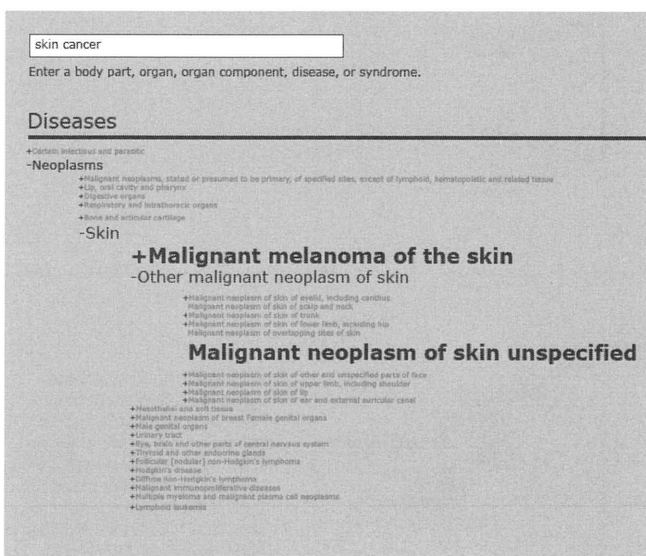

Figure 5. A student designed example using the Fisheye technique. Designer: Weston Morris

Based on advances in the science of perception, and the work of Colin Ware and others, perceptually based theories have been advanced to facilitate information visualization (Ware, 2004). In particular, the parameters governing pre-attentively processed visual form have been shown to have a positive relationship to information that 'pops-out' from its surrounding. Designers are studying pre-attentively processed form. For example, one study defined the pre-attentively-processed feature of blurriness as a ratio of solid to tinted pixels *(figure 6-7).* The ratio at which the blur was immediately distinguished was established. It was noted that lighter value objects, having less edge contrast, required more blur than darker, higher edge contrast objects, to be equally pre-attentive. Establishing ratios that have a proven effect could help visualization designers control how strongly various parts of a visualization appear. The control of hierarchy is essential for effective visual communication. While much remains to be done to define the parameters of pre-attentively processed visual form, it is hoped that design principles based on it can be used by designers to create computer mediated systems where the data, the computer and a user can cause various and alternate data features to pop out based on user interactions.

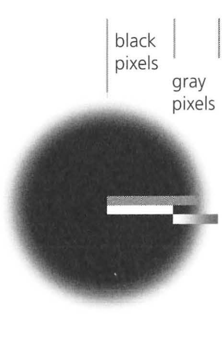

Figure 6. Blur was defined as the ratio of black (solid) to gray (tinted) pixels. Research designer: Chrissie Talkington

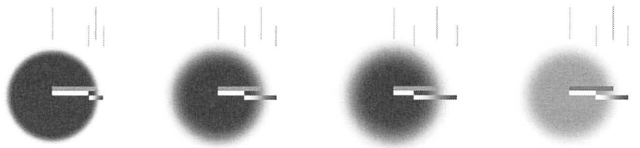

Figure 7. Various ratios were tested for pre-attentiveness (recognized immediately, technically under 10 msec). The gray lines indicate the circle size prior to blurring; the white lines indicate the number of black (solid) pixels; the gradated lines indicate the number of gray (tint) pixels. A ratio of about 1:1.2 and above was determined to be pre-attentively processed when the circles were black. A higher ratio, greater amount of blur, was needed for gray to appear equally burred within a field of sharp and blurred gray dots.

Other studies have explored means of expanding the communication potential of glyphs and icons (Zender, 2006). For icons, the role of context is as relevant to non-verbal communication as it is to verbal communication. Three levels of context were defined and shown to have promise for refining communication with icons. Other manipulations of context have shown potential, such as degrees of abstraction and icon sequences. In addition to icons, glyphs, which are non-representational visual form systems, have promise for visualizing multivariate data (Ware, 2004).

FACETS OF A SOLUTION

Together, developments in controlled vocabularies, information visualization and the design of visual icons provide a foundation for representing medical concepts published in papers. A key remaining obstacle to designing a prototype solution is the scope of the problem. Medicine, even the UMLS, is so broad as to be difficult to effectively design and test. In order to make the problem tractable, a large but defined domain was selected. We hypothesized that demonstration of the feasibility of this approach in one area should be extensible to other areas of science. As a test case for graphic visualization, we focused on how to visualize what is known about a large knowledge domain that is defined yet incompletely understood: the etiology of Alzheimer's

Disease. Another remaining obstacle is the paucity of collaboration between science and design. These two fields, though related in problem-solving methodology and complementary in concern for bringing understanding to data, seldom work collaboratively. To address this obstacle the authors formed an interdisciplinary collaboration. One of us (KAC) works in the field of Alzheimer's Disease research. The other one of us (MZ) works in the areas of digital visualization and non-verbal communication.

Having selected a focused domain and having the necessary expertise to guide the project, we sought to represent key biological/medical concepts associated with Alzheimer's Disease using glyphs and/or icons. Key concepts are defined as those that are essential to describe a hypothesis or experimental finding.

The general approach was to identify key concepts, connect those concepts in summary statements, break those statements into their essential conceptual objects, illustrate those concepts using icons and glyphs and present these visual objects in an interactive concept space where they can be immediately perceived and understood in relation to each other. The perception of concepts in context is expected to facilitate exploration and discovery.

A key problem at the outset was how to extract propositions from published papers. For our project, papers were reviewed manually based upon a random selection of forty papers from PubMed based on a search with the terms 'Alzheimer's Disease' and the protein 'ApoE' (one area of Alzheimer's disease research with which one of us [KAC] is familiar). From these papers, twenty propositions were extracted that express key concepts. These statements *(figure 8)* are stated in positive and negative terms in order to expand their meaning to include both sides of the proposition.

With advances in Natural Language Processing (NLP) and related techniques, the process of extracting concepts from published papers should eventually become automated. Significant progress has been made in this area by others (Revere, 2003) and demonstrated in the Telemakus system (www.telemakus.org). Our study focused on the visualization and display of concepts rather than their extraction from the literature. In the final analysis, extraction of data is of little or no use in solving the problems stated in the introduction to this paper

1	Polymorphisms of apoE are (not) associated with the risk of Alzheimer's disease.
2	Polymorphisms of apoE are (not) associated with the risk of multiple sclerosis.
3	Polymorphisms of apoE are (not) associated with the risk of autism.
4	Polymorphisms of apoE are (not) associated with glaucoma.
5	Polymorphisms of apoE are (not) associated with outcomes following head injury.
6	Polymorphisms of cathepsin D are (not) associated with the risk of Alzheimer's disease.
7	Proteolysis of apoE is (not) associated with neuronal degeneration.
8	ApoE does (not) regulate metabolism of ß-amyloid.
9	Estrogen does (not) modulate the expression of apoE.
10	Polymorphisms of apoE do (not) interact with herpes simplex virus to modify the risk of Alzheimer's disease.
11	Cathepsin D does (not) degrade apoE.
12	ß-amyloid does (not) cause neuronal degeneration.
13	The C-terminal fragment of apoE does (not) bind to ß-amyloid.
14	ApoE is (not) required for plaque formation.
15	Cathepsin D is (not) present in plaques.
16	ApoE is (not) present in plaques.
17	ApoE is (not) produced by macrophages.
18	ApoE is (not) produced by nerve cells.
19	ApoE is (not) produced by astrocytes.
20	ApoE does (not) affect long term poteniation (LTP).

Figure 8. Propositions (or hypotheses) from papers

unless the data are effectively presented. This has been demonstrated repeatedly in Edward Tufte's seminal works, particularly with the story of the 1986 space shuttle Challenger disaster (Tufte, 1997).

The proposition statements were then categorized roughly in line with the UMLS and with the relevant MeSH terms *(figure 9)*. We referred to these categories as 'objects' in the sense of modular conceptual elements, like individual words, that could easily be rearranged to make statements. Such objects fall broadly into things, actions and modifiers, analogous to nouns, verbs and adjectives/adverbs in language. One interesting challenge is to communicate processes as objects. Such objects are roughly similar to gerunds, or verbal nouns, in language. Gerunds in English are generally formed by adding 'ing' to the end of a word, such as 'time'. 'Timing' is a noun that denotes a process (verb) by adding "ing" to "time." In the propositional statements we analyzed, "neuronal degeneration" is one example of a process object: a neuron (thing, noun) degenerates (dies). This would be a conceptual entity in the UMLS.

Following the identification of the necessary objects and their conceptual categories, student designers, working under the direction of the authors, converted each conceptual object into a visual icon/glyph. The icon/glyphs were conceived not as isolated visual objects but as an integrated system of communication objects designed to be read together. Designing icons to work together adds to the context of the entire system so that each icon helps inform the interpretation of every other icon. The role of Proximate Context, the field of interaction where images in a system interact with other images in the same system, has been described elsewhere (Zender, 2006). The ultimate aim was to combine icons with more abstract visual shapes and icon modifiers in a system that could express complex visual concepts; one such system is illustrated here *(figure 10)*.

BIOLOGICAL OR BIOCHEMICAL OBJECTS
polymorphisms of apoE / apoE
C-terminal fragment or apoE
cathepsin D
ß-amyloid
estrogen
herpes simplex virus
plaques
macrophages
nerve cells
astrocytes
PROCESS OBJECTS
proteolysis
neuronal degeneration
long term potentiation (LTP)
DSEASE OBJECTS
Alzheimer's disease
multiple sclerosis
autism
glaucoma
head injury
herpes
ACTIONS
does regulate / does not regulate
does modulate expression / does not modulate expression
does interact with / does not interact with
does cause / does not cause
does bind / does not bind
is associated / is not associated
is required / is not required
is present / is not present
is produced by / is not produced by
does affect / does not affect
does degrade / does not degrade
to modify

Figure 9. Objects found in extracted propositions

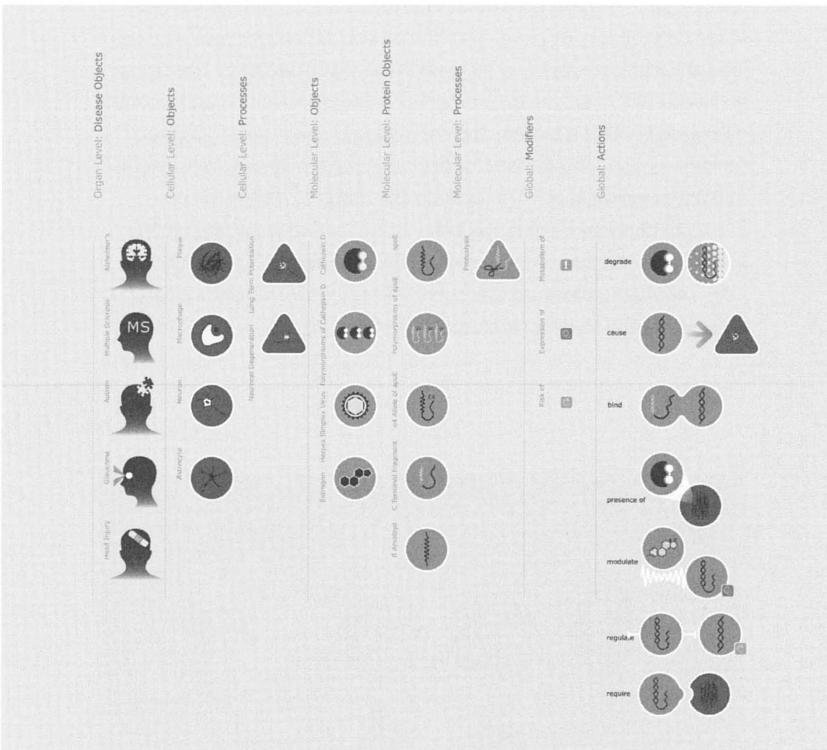

Figure 10. Icon system to express medical concepts design: Sean Gresens, David Kroner, Nolan Stover and Luke Woods

This system is organized by physiological levels moving (left to right) from the higher level to lower: organism (in this system human – not illustrated), organ, tissue, cellular and molecular levels. The disease objects are represented at the organ level by a human head in very dark purple. The darkness and complexity of the contour of the head shape differentiate this level from the others. At the next level, the cellular, the icons are enclosed by medium dark purple circles. The cellular and molecular levels are in turn distinguished by medium dark and light purple circles. In this system, the containing shape is serving a glyph function: a non-representational form used systematically to distinguish one physiological level from another. The ability of glyphs to communicate multivariate data has been well documented (Ware, 2004). In the system shown, the containing shape serving the glyph function adds important context to the icons. An additional containing shape, a triangle, is used to designate processes *(figure 11).*

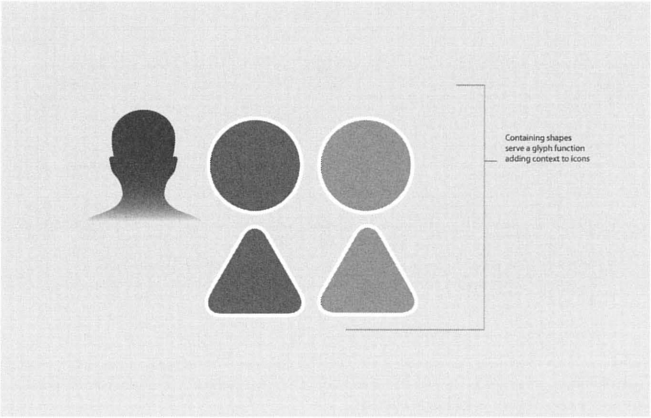

Figure 11. Containing shape serves glyph function

In addition to containing shape glyphs, the icons themselves were designed using a continuum of detailed organic form to geometric form to support reading of icons at the proper physiological level. More organic shapes, those having less regular contours, were used consistently for higher physiological levels while more geometric or regular shapes were used for cellular and molecular levels *(figure 12).*

The visual representation of physiological levels in this example was deemed important not only for scientists' conception of the organization of information but also to provide sufficient context to interpret the icon meanings. The role of context in clarifying meaning has been explored previously (Zender, 2006). As noted above, context informs meaning in everyday experience as well as in written language. The UMLS accounts for the role of context in its use of specialized vocabularies that assign the correct meaning to a word by placing it in its

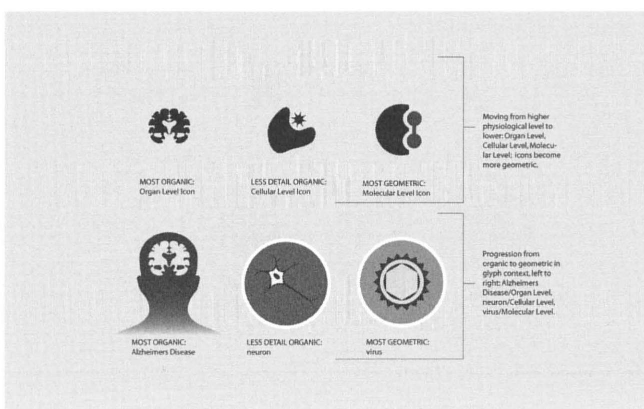

Figure 12. Icon shape contour: Organic to geometric, contributes to understanding

relevant vocabulary context. The same principle holds for visual language where an icon's meaning is influenced by its context. In the system shown here *(figure 13),* the globular form seen in the context of the cellular level was interpreted correctly by a viewer as a macrophage, whereas a similar globular form in the context of the molecular level was correctly interpreted as a protease (cathepsin D).

In addition to using containing shape as a glyph, this system also uses three signs as modifiers: 'metabolism of,' 'expression of' and 'risk of' (top to bottom). The meaning of these symbols must be learned, as all symbols must be, because they have no direct representational connection to the concepts they express.

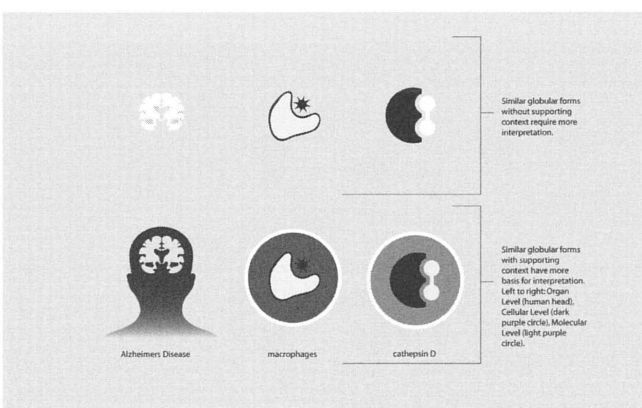

Figure 13. Globular forms clarified by glyph context

Other icon studies have used more suggestive visual form as modifiers, e.g., glow for heat, or scribble for negation (Zender, 2006). An orange glow, even in a very abstract form, is more representational than an arbitrary symbol such as an exclamation point. An orange glow in the right context can suggest the concept of heat, and in the context of a more representational icon such as an icon of a hot dog, a glow might mean warm hot dog. The glow just described is intermediately abstract: it suggests a representational experience without representing a specific recognizable object. Using higher degrees of abstraction to represent higher conceptual concepts and lower degrees of abstraction to represent more tangible objects has been explored by one of the authors (Zender, 2006). In this system example, the icon for 'degrade' uses an intermediately abstract series of dots to eat away at the icon, similar to corrosion of metal *(see figure 14 below).*

At each organ level glyph, a human head for example, are one or more additional icons to represent different diseases. For example, a bandage icon is placed on the head icon to represent head injury (trauma). Although this system generally adds only one icon to the head shape to specify meaning, the 'glaucoma' icon suggests that more than one icon could be imposed on the head to build more specific meaning. Other icon systems developed by other student teams more fully demonstrated this possibility. Note that with one exception (the head icon containing the letters 'MS' representing multiple sclerosis), none of the icons relies on letters or words, though words were added later as rollover 'tool tips' to enhance clarity and speed learning.

The icon system shown here makes effective use of Proximate Context: one icon in the system informing the interpretation of other icons in the system. For example, at the molecular level the icon for 'apoE' *(figure 10, top)* is repeated with modification to represent 'polmorphisms of apoE' and 'the E4 allele of apoE.' A fragment of the 'apoE' icon: the zigzag form, is used elsewhere to represent other protein concepts such as 'C terminal fragment' and 'proteolysis.' The same zigzag portion of the 'apoE' icon is used alone to represent a different protein fragment, 'beta amyloid.' A similar pattern exists at the cellular level with the 'neuron' icon and the process icon used to represent 'neuronal degeneration.'

As noted above, the objects visualized include both nouns and verbs. Visualizing actions (verbs) is a significant challenge yet essential for representing the propositions above. Existing icon systems typically do not have icons for verbs (Zender, 2006). One approach to visualizing an action is to combine a physical icon with a separate action icon or glyph. The action objects are represented in the system example by icon/glyph combinations that were created in both static and animated forms. The static ones shown below *(figure 14)* in some cases apply a glyph modifier to two existing icons to suggest an action, e.g., dots 'degrade' a protein icon. In other cases the surrounding shape is modified to suggest the action: the C terminal containing shape and protein containing shape are fused to represent 'binding.' This is an inconsistent solution but, in the limited scope of this system, it functions reasonably well. As noted above, the static versions of the actions shown here also had animated versions, making the meaning more explicit. The use of animation to clarify iconic and glyph-based visual communication has not been widely explored but shows great potential (Ware, 2004). With the common use of computers for both creating and displaying information, animation can be used to express data and is relatively inexpensive to develop.

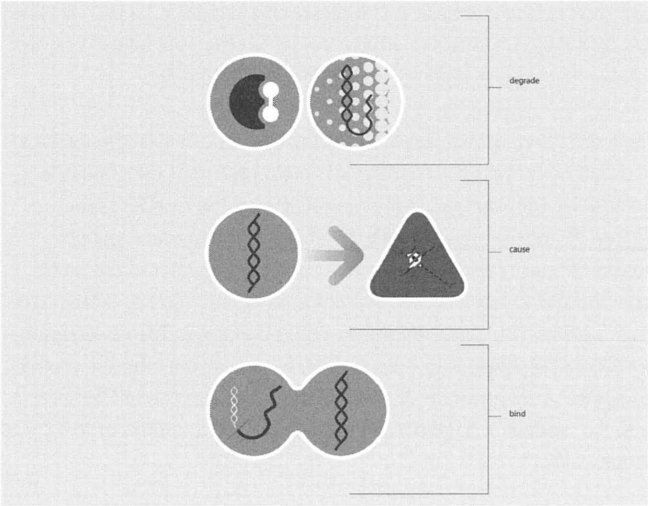

Figure 14. Action objects

The icon systems developed were not ends in themselves but were designed to be composed into statements representing the original propositions extracted from the papers extracted from PubMed as the basis for this work. In the icon system shown here, icons were tested as combinations that could be read as propositions. An example of two such icon proposition statements are shown in Figure 15. In informal testing, subjects were able to quickly read each proposition accurately from the icon/glyph presentation after minimal experience – one to two minutes - with the iconic system.

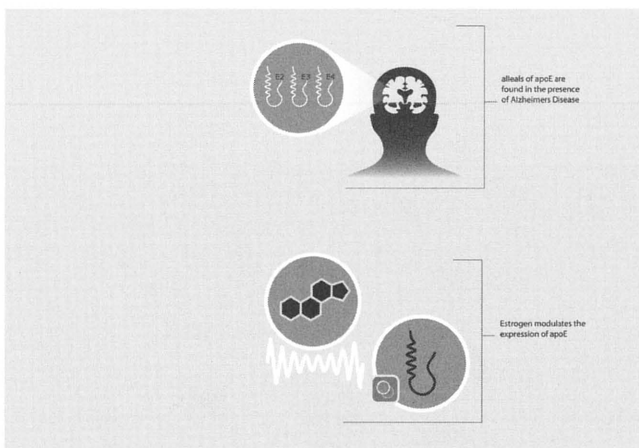

Figure 15. Icon proposition statements

The final step was to combine visual propositions like those shown above into an interactive display. The display example shown is an interactive prototype only, not driven by a database. The intention for the final display is for the icons and other graphic elements in the system to be stored in the database with tags to the data and displayed in response to a query. In the prototype shown below *(figure 16)* this is simulated by typing in 'apoE.' The result is a cluster of icons in various sizes. The center icon is the search term, surrounded by associated concepts.

Surrounding icons are clustered loosely into zones reflecting the organizational levels: disease in the lower right moving clockwise through cellular and molecular levels. The smaller the surrounding icon and the further from the central icon, the less support there is for the association (quantity of papers in PubMed or some other relevant measure). Roll over each icon and a tool tip identifier appears. Click on a concept pair and a window displays all the relevant papers that report on the concepts. The concepts can be sorted by function using a pull-down menu. Sorting the display by the concept 'association' blurs the concepts that do not have 'associative' relationship. Similarly, the display can be resorted by clicking on any icon, which launches a new search, pulls in additional icons and reorganizes the display with the newly selected search term in the center.

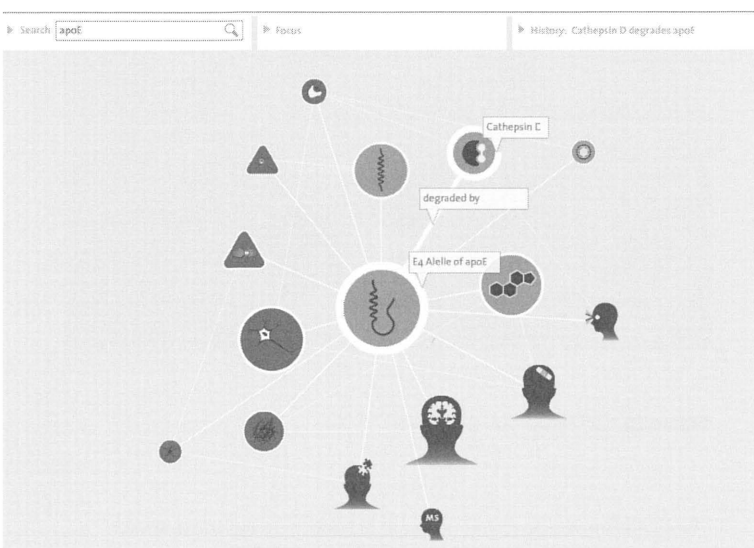

Figure 16. Interactive information design – Alzheimer's display

There are many more possibilities yet to be explored with this display approach and it is clear that development of an optimal system will require an interative process. The example shown above has been informally tested with scientists with positive results. Most are able to quickly read the concepts and find the interaction to be both intuitive and informative.

Failings of the example shown include the failure to display the animated verbs in the original icon display. This reduces the effectiveness considerably. The display has not been tested in direct comparison with similar displays presented as words. Further testing is definitely needed, even though the results to date are promising.

SUMMARY

You can see a lot. And as we have suggested in the introduction of this paper, you can see a lot more quickly than you can read. We believe that visualization holds great potential, not to replace reading, but to summarize verbal content. In areas with controlled vocabularies and clear ontologies, we believe icon/glyph systems may be developed that effectively communicate with their constituents. By making such visualizations interactive they have the potential to move beyond description to become tools of discovery. Simpler graphic means have been proven to do so (Bertin, 1989), why not more explicit means? If such visualization techniques are developed it may be possible to expand the dialogue of science beyond the verbal realm into the visually symbolic realm. We may discover that cave paintings and tomb hieroglyphs were a good idea awaiting the development of technology that can generate icons and glyphs rapidly, interactively and inexpensively.

REFERENCES

Arnheim, Rudolf. 1969. *Visual Thinking.* Berkley: University of California Press.

Arnheim, Rudolf. 1974. *Art and Visual Perception.* Berkley: University of California Press.

Bertin, Jacques. 1981. *Graphics and Graphic Information Processing.* Berlin: Walter de Gruyter & Co.

Card, Stuart, Jock D. Mackinlay and Ben Shneiderman. 1999. *Information Visualization Using Vision to Think.* San Francisco: Morgan Kaufman.

DaFanti, Thomas A., Maxine Brown and Bruce H. McCormick. 1987. *Visualization,* NSF Report, Computer Graphics / IEEE Computer.

Furnas, G. W. 1981. *The FISHEYE view: a new look at structured files.* Murray Hill, NJ: Bell Laboratories.

Lyman, Peter, Hal R. Varian. 2000. How Much Information, report, University of California at Berkeley, School of Information Management and Systems. (see also 2003 edition) excerpt from http://www.press.umich.edu/jep/06-02/lyman.html

Lupton, E. 1989. *Reading Isotype. In Design Discourse: History/Theory/Criticism.* Chicago: University of Chicago Press.

Revere, Debra. 2003. A New System to Support Knowledge Discovery: Telemakus. In *Proceedings of the American Society for Information Science and Technology Annual Meeting.*

Spence, Robert. 2001. *Information Visualization.* Harlow, England: Addison-Wesley.

Tufte, Edward. 1983. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press,

Tufte, Edward. 1997. *Visual Explanations.* Cheshire, CT: Graphics Press.

Wright, N. T. 1992. *The New Testament and the People of God.* Minneapolis: Fortress Press.

Ware, Colin. 2004. *Information Visualization: Perception for Design.* San Francisco: Morgan Kaufman.

Zender, Mike. 2006. Icon Systems for Global Non Verbal Communication. *Visible Language* 40.2, 177-206.

ACKNOWLEDGEMENTS
Thanks go to the Junior Digital Design student team who designed the icon system used as the example in this paper: Sean Gresens, David Kroner, Nolan Stover and Luke Woods. Thanks also to Chrissie Talkington who developed the study on blur and Weston Morris who designed the Fisheye example *(figure 5).* Thanks also to the other students and faculty at the University of Cincinnati College of Design Architecture Art and Planning who over the years have participated in discussions and contributed alternative design approaches that necessarily inform the approaches described in this paper.

AUTHOR NOTES

MIKE ZENDER is a third generation designer, following in the footsteps of his grandfather, father and now teaching and working with his son Micah. Mike founded the firm Zender + Associates Inc. after attending the University of Cincinnati and Yale University. He currently teaches digital design at the University of Cincinnati with teaching and research emphasis in information design, visualization and informatics, including recent studies in icon systems, glyph systems and their role in communication and visualization.

KEITH A. CRUTCHER, Ph.D. is a professor in the University of Cincinnati's College of Medicine and a neuroscientist with active research programs in Alzheimer's disease and neuronal plasticity. He has co-authored several scientific papers and deals daily with the problem of how to cope with the explosion of relevant literature in his areas of research.