Proceedings of Human Factors and Ergonomics Society, 2005, pp.1561-1564

ON REDUNDANCY IN THE DESIGN OF SPATIAL INSTRUMENTS

Stephen R. Ellis
NASA Ames Research Center
Moffett Field, California USA

Alternative formats for information displays are sometimes identified as being best suited for presentation of particular types of information. This view is assessed in terms of the role of noise and distortion in the presentation of spatial information. It is shown that introduction of redundant elements may compensate for weaknesses in different formats. Consequently, it is argued that the observed differences among formats may in fact arise from specific design decisions relating to the redundancy in the presented information rather than features inherent to the formats themselves.

INFORMATION AND MEANING

There is no information in the physical universe. Information only arises in the presence of an objectively established intentionality that establishes the existence of a signal. A signal is a desired time-varying pattern of energy associated with some goal or aim belonging to an agent who/that inhabits the signal's environment. Signals may be either continuous or discrete but today are generally discrete, digital codes.

In this context an information display is a communications medium over which the signal is sent to a human user. Historically, displays have been mainly unidirectional like pictures or sign posts. But since the Industrial Revolution, and more intensively since the widespread use of computer-based devices, displays have become increasingly interactive. Hence, users both receive and send signals via displays.

The purpose of a display, however, is not the transmission of information but of meaning. Information transmission is only a necessary but not a sufficient condition for a successful display. The meaning is provided by the semantic context surrounding the receiver and the sender through their knowledge of other signals with which the signal is semantically and syntactically associated. Because the measurement of the quantity of information in a signal is determined by its "surprise value" or unexpectedness[1], its information content, though important, is not necessarily closely related to its meaning.

Meaning, in fact, is a kind of dual of information. When a signal or message has a significant semantic context, the transmission of subsequent specific signals becomes more expected and their information content is consequently reduced by the redundancy[2]. English text, for example, has been argued to convey 0.6 – 1.3 bits/character for a 27 character alphabet[3] (Shannon, 1948). This value corresponds to a redundancy of 72-87% since maximum information per character would be 4.75 bits/character. Some of this redundancy may be captured in a purely statistical Markov process, but syntactic and semantic approaches are required for a more complete computational model.

Figure 1 provides several examples of what this high level of purely statistical redundancy means in terms of text showing several orders of statistical approximation of English. A zero order approximation is based only on the probabilities of the appearance of each character. A first order approximation is based on the conditional probabilities for a character given a previous character. A second order approximation is based on probabilities conditioned on preceding character pairs, etc. We can see that by the third order approximation, the stochastic processes begins to generate a number of recognizable English and English-sounding words, though no meaningful message emerges. The point here is that user-system communication of all types needs to have considerable redundancy to match the kind of communication which users habitually process.

0 – order letter

xfoml rxkhrjffjuj zlpwcfwkcyj ffjeyvkcqsghyd qpaamkbzaacibzlhjqd mhjjw

2nd – order letter

on ie antsoutinys are t inctore st be s deamy achin d ilonasive tucoowe at teasonare fuso tizin andy tobe seace ctisbe.

3rd – order letter

in no ist lat whey cratict froure birs grocid pondenome of demonstures of the reptagin is regoactiona of cre.

Figure 1. Statistical approximations of English based on the probability and the conditional probability of appearance of letters in text (Shannon, 1951).

The amount of redundancy required for successful communication is fundamentally determined by the statistics of the message and of the sources of noise which set a limit to the rate information may be transmitted over a channel (Shannon,

---

[1] Information, H, measured in bits and associated with a set of received signals(symbols) $a_1, a_2, a_3, \dots a_n$ which have the probability of appearing in the ensemble of received signals of $p(a_1), p(a_2), p(a_3), \dots p(a_n)$, is $H = \Sigma_i -p_i log_2(p_i)$

[2] Redundancy, R, for a signal having information content of H may be expressed as a percentage of the maximum information content it could carry H_{max} which corresponds to the situation in which all the elements of the set of all possible signals that could be received have equal probability, $R = ((H_{max} - H)/ H_{max})*100$

[3] A space is also a character.

1948; Cherry, 1978). Optimal use of a communications channel thus requires a matching to the interpretative capacities of the users. Appropriate coding schemes can aspire to maximize information rates (Mandelbrot, 1982), but designers have to be careful not to over do it. Insufficient redundancy can make the semantic content susceptible to degradation of noise and distortion.

SPATIAL DISPLAYS

Spatial displays transmit information and meaning just as do text messages. The meaningful characteristics of a display, particularly a display involving spatial elements, may be classified into three categories: geometric, dynamic, or symbolic. The geometric features are those describing position, orientation, adjacency, proximity, and connectedness, i.e., the classic geometric characteristics. The rules governing change in the display, e.g., velocities and accelerations, as well as in state changes such as color in visual displays or timbre in acoustic displays, are the dynamic characteristics. Those features of the display elements that obtain independently of the element's position or state of motion/change are the display's symbolic features. They could include such static characteristics as shape, smoothness, roughness, etc. Significantly, these elements may also have their own internal dynamics, e.g. rules for temporal changes of shape.

Breaking the display features down into these three categories is not just an academic activity. Each feature and associated subfeatures provide a channel that may be used to communicate information and meaning to a user. Because of variations in the transmission environments, e.g., the meaning or context of the intended "messages," and the physical properties of the human sensing systems, communication along any one of these possible channels will have definite limitations. The challenge for any display designer is to insure that sufficient capacity is available for the specific messages and signals they wish to send. This involves matching the coding system to each channel to optimize its use (e.g., Mandelbrot, 1982) but it also can involve cross feature enhancements. The geometric, dynamic, and symbolic features can be mutually supporting and thereby provide an increased channel capacity and signal redundancy. Designers should think broadly when considering such interfeature support since it may not only involve different sensory modalities such as vision, audition, or haptics, but can involve within mode enhancements of the geometric, dynamic, or symbolic features. Some examples of such enhancements are provided below.

SPATIAL DISPLAYS AND SPATIAL INSTRUMENTS

Perspective displays are widely known to introduce apparent spatial compression into the 2D images on their projection planes. But carefully matched wide-angle distortion can be exploited to compensate for the compression that would otherwise be present on a "correct" projection viewed from the geometric center of projection (McGreevy & Ellis, 1986; Grunwald, Ellis, & Smith, 1988). This design feature presents truth through distortion much as cartographers do by accepting map distortion of some features for accurate presentation of others. This accepted distortion is an example of geometric enhancement. The introduction of such an enhancement with the goal of supporting specific communicative needs turns a spatial display into a spatial instrument.

Displays in modalities other than vision may be similarly enhanced. Force-reflecting haptic displays are known to develop task-limiting instabilities in the presence of time lags[4] (e.g., Kim, Hannaford, & Bejczy, 1992). Designers of such telerobotic systems have learned that introducing some compliance, i.e. springiness, either into the mechanical linkage or the control software can resolve this problem. Because such a design modifies the way forces and torques are propagated through the system and displayed to the user, it may be described as a dynamic enhancement.

An example of a symbolic enhancement may be found in an audio display: When the physical properties of sounds are kept constant, sounds with meaningful associations are more readily detected against a background of noise (Miller, Heise, & Lichten, 1951). Use of sounds with symbolic associations can by itself thus improve effective channel capacity

INFORMATION REDUNDANCY

Successful transmission of a message can be blocked by random or systematic processes which respectively correspond to noise and distortion. The following examples illustrate how the addition of graphic redundancy to visual spatial displays can combat these two types of disturbance.

Noise

Tufte (1983) has extensively discussed techniques for the visual presentation of quantitative information using a technique of comparative analysis akin to approaches in art and design classes. With a goal of trying to increase the amount of information that a given amount of "ink" can convey, Tufte calls one of the metrics he has identified the data-ink ratio[5]. The three panels of Figure 2 illustrate how the amount of "ink" devoted to a graph of the same data can be reduced with the amount of specific information provided by the remaining 'ink" actually increasing. As can be seen, though the graph at the far right might actually convey more information including, for example, the actual data values, it begins to lose visual coherence as the data points and the numbers representing the coordinates of specific points appear to mix visually, making it harder to see structure in the data. In this case the greater redundancy and reduced data-ink ratio in the graph at the left better stands up to added visual noise as illustrated in Figure 3. It demonstrates that, just as with text messages sent over a

---

[4] This threshold for jitter or instability strongly depends on criteria for task success and varies widely with citations from 30 to 500 msec.

[5] The data-ink ratio = (ink dedicated to non-redundant presentation of data)/ (total ink used to print the graphic).

noisy channel, the redundancy of visual signals needs to be controlled to insure their receipt.

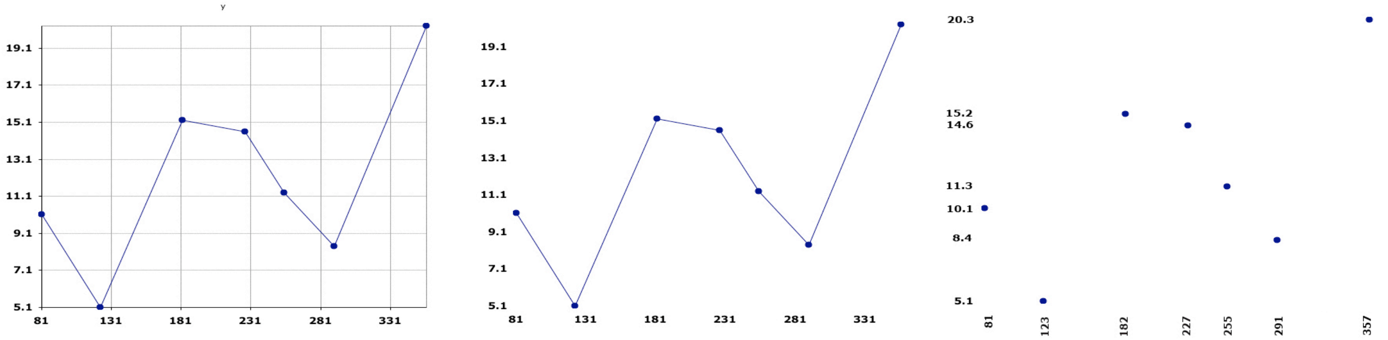

Figure 2. Three plots of the same data with decreasing graphical redundancy (Tufte, 1983).

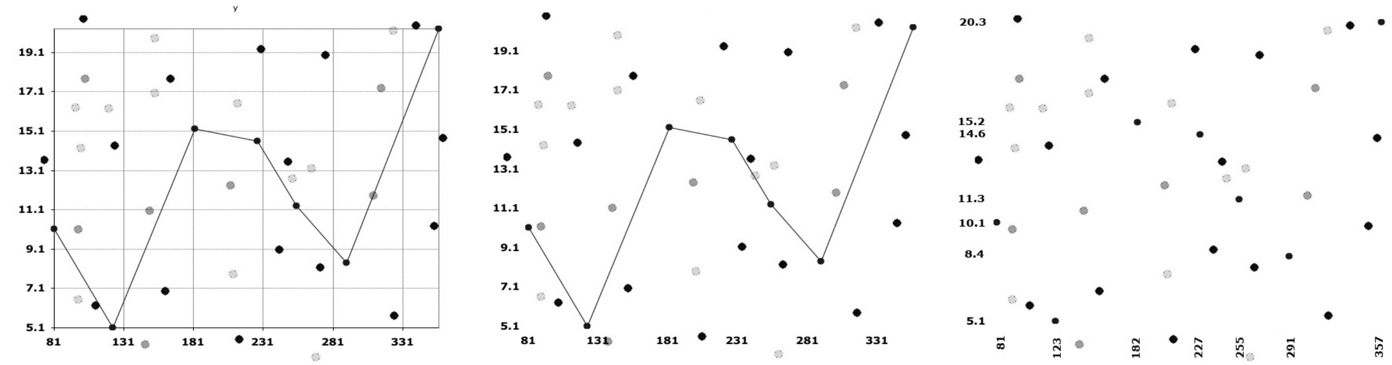

Figure 3. Identical random visual noise is added to the panels of Figure 2 illustrating how the graphical signal in the least redundant version at the right of the figure is lost. The additional display elements of the leftmost panel are redundant symbolic enhancements assisting communication in the presence of noise.

Distortion

In contrast to the effects of noise, some constraints on visual communication result from systematic processes. One example is the problem of pose ambiguity present in perspective images as shown in Figure 4 (Smallman, St. John, Oonk, & Cowen, 2001). This is a distortion that occurs in an insufficiently redundant spatial instrument presenting three-dimensional data.

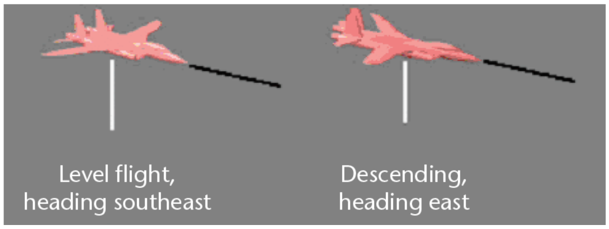

Figure 4. Pose ambiguity illustrated by two realistic aircraft icons in different true poses but with very similar projected images .

Pose ambiguity is well known in M.C. Escher's artwork in which he exploited it for artistic effect, but it also has been studied in the context of air traffic information displays (Ellis, 1993). Aircraft icons, such as those introduced in Figure 4, can be modified, (Figure 5), so as to minimize instability or ambiguity in their pose through the introduction of orthogonal visual contours. These contours help stabilize perception. They, however, must be visually close to the objects with which they interact. As shown in Figure 5, an aircraft icon raised above its stabilizing ground grid, can appear to rotate into the display surface as its aspect with respect to the perspective projection becomes less and less informative. What is happening is a kind of regression to the projection surface as the pictorial cues to the aircraft's true pose grow visually more distant and weaker. The addition of a second reference line at the end of the predictor of the aircraft's future position helps stabilize and clarify the apparent pose of the icon as illustrated in the lower part of Figure 5.

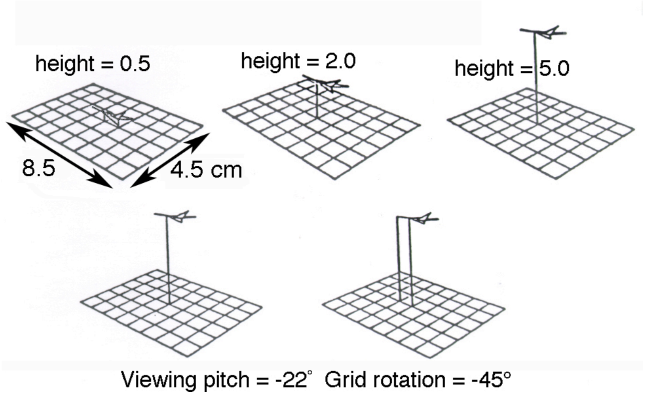

Figure 5. Five views of sample stimuli used to examine the perceptual effect of raising an aircraft symbol above a reference grid. The attitude of the symbol was kept constant relative to the grid and viewing cameras as

it was raised to different heights above the grid. The lower panels illustrate the effect of the addition of a second vertical reference line from the extended "nose" of the symbol. It reduces the illusory rotation of the aircraft relative to the grid caused by increasing the height.

The impact of this technique of a second reference line to reduce pose ambiguity has been quantified in an experiment. The results are plotted in Figure 6. In this experiment ten naive subjects aged 18 to 45 viewed 54 distinct static perspective projections of aircraft-like symbols elevated at three different levels above a ground reference grid: a low level below the axis of the viewing vector, a middle level in which the viewing vector pointed directly at the aircraft position, and a high level for which the aircraft was above the viewing axis. The aircraft symbols had straight predictor vectors projecting forward, showing future position. In one condition, reference lines were dropped only from the current aircraft position; in the second condition, redundant lines were dropped from both current and predicted position.

Nine different grid rotations presented in equal 22.5° intervals from 0 to 180° were crossed with the three aircraft heights and the presence or absence of the second reference line in a fully factorial repeated measures design making the 54 distinct conditions. When viewing the images from about 50 cm, subjects responded to each by adjusting a pointer on a protractor that was flat on the table supporitng a mount that held printed images of the display perpendicular to the subjects' view. A research assistant randomized the order of presentation, manually advanced the images, and recorded the subjects' orientation judgment. Since the phenomenon had been easily seen in pilot experiments, there were no replications within subjects. It is important to note that the subjects needed minimal practice, making only one or two practice judgments before the experiment started. However, it was made clear to the subjects that they were to respond to the egocentric orientation of the aircraft symbol, i.e. the orientation with respect to their body. They were also told that it could be in any possible orientation and that they should try to determine it as accurately as possible. As noted in the caption to Figure 6, the three-way interaction of grid orientation, aircraft height, and the number of reference lines was highly statistically significant.

The first observation from the experiment was that though there were no pose reversals, subjects made substantial errors in their estimation of the azimuth rotation of the aircraft; they generally saw it rotated more towards their frontal plane than it in fact was. It seemed to rotate even with respect to the grid it was positioned above. Indeed, some subjects reported that it appeared to twist with respect to the grid. The second result was that the error towards the frontal plane of projection for the symbols with one reference line increased as the height of the symbol increased above the grid. This additional error in apparent azimuth arises as the aircraft symbol is increasingly visually displaced from the ground reference grid which provides useful pictorial cues to the symbol's depth into the image. Most significantly, however, introduction of a redundant second reference line almost totally eliminated the effect of symbol height, reducing the overall azimuth error in some cases by almost 50%. Thus, pose distortion, a supposed difficulty inherent in perspective displays, is shown not to be an inherent feature of this format. The second reference line is, however, not completely redundant with respect to the first line: The second line is neither identical in position nor shape but, because it is attached and orthogonal to the same object as the first line, it is remains spatially correlated with it.

Two elements of a display that are highly correlated in position and shape may indeed be judged highly similar and may consequently be descriptively confused. Confusions of this type have been employed by Bjorke (1996) as part of a framework he has developed for calculating cartographic entropy. Since maps are classic examples of spatial displays, his discussion applies directly to the quantification of the graphic redundancy discussed in this paper. In his 1996 paper he focuses entirely on the syntactic aspects of the entropy, i.e., information content, of maps, but his analysis applies directly also to spatial displays.

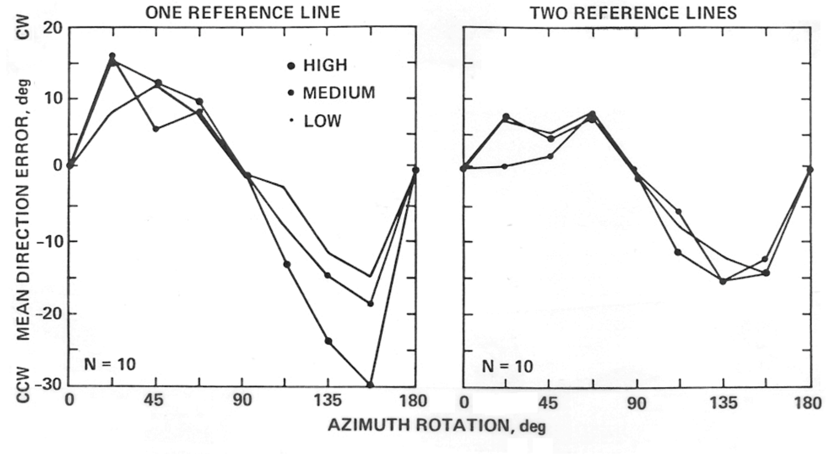

Figure 6. Mean clockwise and counterclockwise egocentric direction judgment errors of the orientation of an aircraft symbol randomly rotated into nine different azimuths. The two panels represent a statistically reliable 3-way interaction from a repeated measures experiment with ten subjects.

(F(16,144)=2.402) p<0.003). The right panel shows that the amplitude of the error was reduced by introduction of a second reference line and that its dependency on the height of the aircraft above the grid was eliminated. Why the effect seems most prominent for the larger rotations is yet to be explained.

A syntactic analysis of the spatial displays discussed above would only consider the relationships among the signs and symbols employed, primarily their positional relationships. The relationships of the symbols to their meaning, their semantics, and of their relationships to their users or applications, their pragmatics, would be excluded from Bjorke's entropy calculations. Accordingly, Bjorke's analysis applies more to the introduction of visual noise illustrated in Figure 3 than to the use of redundancy to improve the spatial interpretation as shown in Figure 5. Semantic context may well be more important for this latter example.

Bjorke does, however, propose that an information theoretic analysis may be applied to several different map features and levels which can be statistically independent information sources. Following Bertin (1978) he identifies them as position, form, position, orientation, color, texture, value, and size and proposes some computational techniques by which the information content originating from some may be computed. Provided that the variation in the separate information sources is uncorrelated, the total information content of the display or map can be determined by adding the entropies for the separate features.

Bjorke repeatedly cautions that the entropy calculations he suggests must be limited to the syntactic features of the display. The heart of his analysis relies on a computation that converts feature similarities into confusion probabilities (Bjorke, 1996),p. 85) which then may be used to compute the equivocation, $H_y(X)$, associated with an information source. This value then may be used with the entropy in the source itself, which is separately calculated as $H(X)$, to compute the useful information in bits presented by the source R,

$$R = H(X) - H_y(X)$$

Such computations may well be applicable to the displays illustrated in this paper. But as Bjorke himself emphasizes, these calculations turn on similarity estimates that can themselves be influenced by the semantic context in which the symbols appear. The similarity of the reference lines in Figure 5, for example, is strongly influenced by whether or not they are given a 3D interpretation. Accordingly, his information calculations are not yet fully automatable. This form of analysis, however, is likely to prove more and more useful as it is further refined in ways allowing it to be applied to both of the illustrations presented in this paper.

CONCLUSIONS

In conclusion, we can see that the weakness or strength which a particular graphic design exhibits due to its formatting can arise from distortion or noise. Interpretative errors can, however, be managed by careful introduction of redundancy. Because there are a great variety of ways redundancy can be introduced, statements about the suitability of particular formats for particular kinds of information should be treated with caution. Such statements may be more about the failure to design signal redundancy than about the display formats themselves. As even Tufte has remarked at the end of his first book (1983), blanket statements about design may mislead the unwary. Truth may be found in distortion and noise may slay even the best design.

ACKNOWLEDGEMENTS

This work was supported by a grant from NASA Exploration Systems' Space Human Factors Engineering program (131-20-30).

REFERENCES

Bjorke, J.T. (1996) Framework for entropy-based map evaluation. *Cartographic and Geographic Information Systems, 23*, 2, 78-95.

Bertin, J. (1967) *Semiologie graphique* Editions Gauthier-Villars, Paris, trans. William J. Berg; University of Wisconsin Press, Madison, Wisconsin, 1983.

Cherry, C. (1976) *On human communication*, MIT, Boston. pp. 205-206.

Ellis, S. R. (1993) Pictorial communication in *Pictorial communication in virtual and real environments*, Ellis, S.R., Kaiser, M.K., & Grunwald, A. eds., Taylor & Francis, London.

Grunwald, A. J., & Ellis, S. R. & Smith, S. R.(1988) Spatial orientation in pictorial displays. *IEEE Trans. on Systems Man and Cybernetics* , 18, 425-436.

Kim, W. S., Hannaford, B., Bejczy, A. K. (1992) Force reflection and shared compliant control in operating telemanipulators with time delay. IEEE Transaction on Robotics and Automation, 8, 2, 176-185

Mandelbrot, B. (1982) *The fractal geometry of nature*, Freeman, San Francisco. pp. 344-347.

McGreevy, M. W. and Ellis, S. R. (1986) The effect of perspective geometry on judged direction in spatial information instruments. *Human Factors*, 28, 439-456.

Miller, G. A., Heise, G. A., & Lichten, W. (1951) Intelligibility of speech as a function of the context of text materials. *Journal of Experimental Psychology, 41*, 5, 329-335.

Shannon, C.E. (1948) A mathematical theory of communication, *The Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656

Shannon, C.E. (1951) Prediction and entropy in printed English. *Bell System Technical Journal, 30,* 50-64

Smallman, H.S., St. John, Mark, Oonk, H.M., Cowen, M.B. (2001) Information availability in 2D and 3D displays. *IEEE Computer Graphics and Applications,* 51-58

Tufte, E. R. (1983) *The visual display of quantitative information,* Graphics Press, Cheshire, Conn., p. 93.