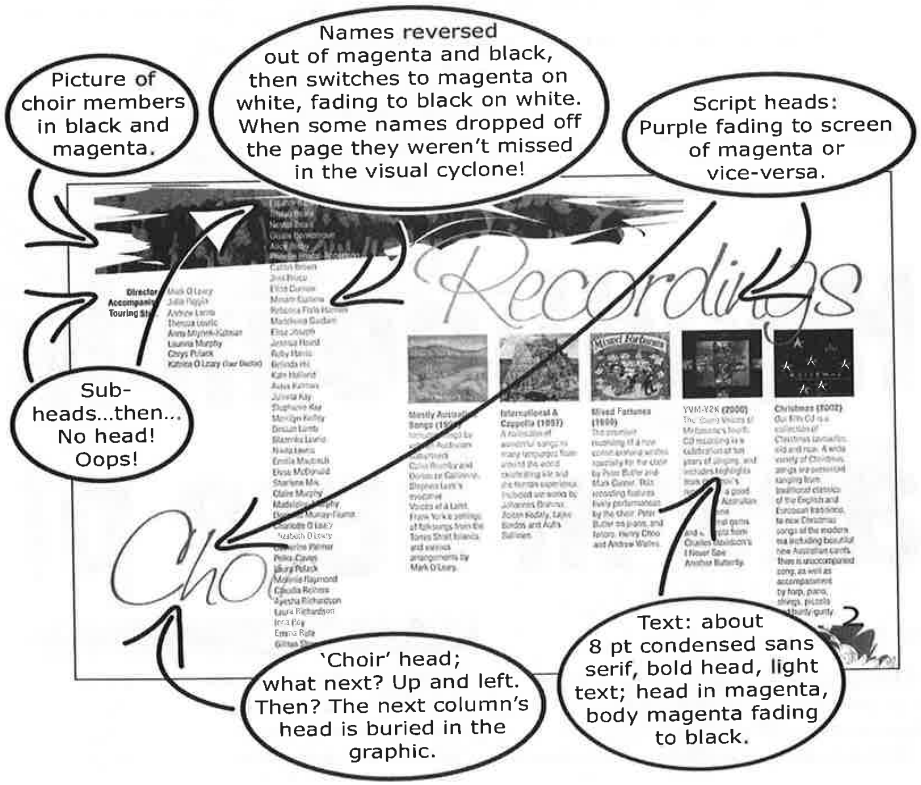
Could granny read it in poor light?

Could granny read this page in poor light — or even at all? That's the question to ask about this children's choir program.

The page looks great, but as reading matter it suffers from half a dozen typographical and layout sins. In the end, one of its major failings is simple lack of consideration for the audience. This is the program for a children's choir. When children's choirs perform, there are lots of grandmothers and grandfathers in the audience and venues are often less than well lit. Combine the typographical nasties with the poor light and smeared granny glasses, and no reading can go on at all.

FIGURE 88: As a design, a piece of art to look at, this spread is delightful; dramatic, full of color and variation! But as a read ...



THE RESEARCH Appendix 1

The Research Program

By Colin Wheildon

My father, a master printer in Derby, England, held three truths to be self-evident:

- Rules and borders must meet like water.
- Serif text is much easier to comprehend than sans serif text.
- Editors and designers are the missing link between the ape world and man.

The first tenet is inarguable.

That it took me a quarter of a century as an editor and designer before I questioned the validity of the second tenet is testimony, in some measure, to the truth of the third.

Early in my career I had become aware that the rules of typography were largely ancient maxims, with very little, if any, documented empirical evidence to support them.

I was aware of research into legibility, which is to say the intrinsic characteristics that make one type easier to read than another.

But I was also aware that research into mere legibility did not provide the answers I wanted.

The title of a newspaper set in Old English text may well be legible, but what would happen if the entire news and advertising content were to be set in Old English? Would it be comprehensible to anyone, other than perhaps a scholar of early English calligraphy or printing?

I had became a disciple of Edmund Arnold, formerly a professor of mass communications at Virginia Commonwealth University, Richmond VA.

FIGURE 89: It may be legible enough as a title, but extended use of Old English type would cause serious comprehension problems.

> In a newspaper nameplate, Old English type like this is appropriate and expected. In the paper's news columns, it would take some getting used to.

Inspired by his common-sense teachings, but alarmed by the fact that even he depended on maxims rather than field research, I determined to subject some of those maxims to research.

This study is the result. It was conducted over a nine-year period, from 1982 to 1990.

The program examined several elements of typographic design. On two of them — the comprehensibility of serif type as opposed to sans serif in body matter, and the comprehensibility of lower case as opposed to capital letter headlines — there is, as I have indicated, some agreement, but very little observation of the agreement.

On three others — the use of color in headlines and in text, and the use of unjustified (ragged) setting, either left or right — there is either great disagreement or complete ignorance of the possible deleterious effects.

On a sixth element examined, whether italic body type is difficult or easy to comprehend, there appears to be a general, but ill-founded, agreement.

Texts on typography frequently allude to research into some of the elements to be examined, but, regrettably, discussion of this research is usually anecdotal rather than empirical.

In an attempt to test the reported findings of overseas research, the major part of this program was carried out in Sydney, Australia, in 1982-86.

About the participants

A total of 224 people, drawn from ten Sydney suburbs, completed the first program — a series of tests run over a period of five years.

The sample initially contained 300 people. As the years passed, some moved to other cities or towns, some died, some dropped out from sheer boredom, and some (I suspect) were committed to institutions for the chronically confused as a result of being bombarded at regular intervals with my experiment papers! All the results papers from those who dropped out were discarded and the responses from the 224 who completed all tests for that period were used for calculating the final results.

Initially, the sample had perfect balance of the sexes, comprising 150 men and 150 women. Fortunately and fortuitously, this balance was nearly maintained, the final figures being 113 men and 111 women.

There were no statistically significant differences between the responses from men and women.

At the beginning of the study, twenty of the readers were high school students. All the adults had completed four years of high school, and 178 (79 per cent) had reached matriculation level (high school graduation). Fifty-two (23 per cent) had acquired a university degree or professional or trade qualification, and 12 (five per cent) had acquired a higher degree. This represented a higher than average education level and obviously was a bias in the sample. This was deliberate. The participants were those who would be expected to take an interest in current events, and consequently would be likely to be regular newspaper and magazine readers.

In fact, all except two indicated they were consistent readers of a range of newspapers and other publications. The two exceptions, teenagers, said they read magazines regularly, but newspapers infrequently.

None of the readers was professionally involved in the printing or publishing industries. All were volunteers.

For the second series of tests, in 1986, the 224 subjects who had completed the first program were called upon again, augmented by a further 276 subjects to make a total of 500. The additional subjects were probably of generally similar background to the first group of subjects, although no attempt was made to identify their interests, education level or to classify them by socioeconomic background. Since this study was of short duration, there were no dropouts so the responses of all 500 subjects were incorporated into the final results.

The flyer study was carried out under the auspices of the National Roads and Motorists' Association (NRMA) in New South Wales beginning in 1988. The 224 original subjects from the first and second programs were again used for the initial comprehension tests. Thus alerted, the NRMA backed a major study conducted through its Public Relations and Research departments. The 15,000 participants were a random selection of members who were randomly assigned to subgroups as required for the different parts of the program. The 4,000 sub-sample used for the type sizes tests was randomly drawn from the 15,000.

Repeated measures

The question of repeated measurements of the same group of subjects has been raised by critics of my results.

Suffice to say that this procedure was approved by my academic and industry advisers who were very experienced in designing and carrying out valid research in the media and other disciplines.

As discussed below, the subjects were divided into two groups for most tests to provide for a control group or a contrasting group, the groups were switched from one parameter to the other, and over the period of five years, the subjects were returned to the total subject pool and reselected into test and control samples time and again.

A further consideration was that there was no logical connection between the different variables being tested. Variables like good reading gravity, type face and type style are all quite separate from each other so there is no need to change samples to measure their effects.

Finally, it ought to be noted that serif type, printed black on white, justified, with good reading gravity, was a variable or the control/baseline in test after test. Doubters should review the results for themselves — across all these tests involving different mixes of the subject pool and different texts, the findings for this combination remained rock solid.

Those with a different view are welcome to replicate my work!

The methodology

The methodology described is that which applied to the first program of the study (from 1982-85). Where it varied in the second and third programs, the difference has been noted (for example, the headline study was not based on comprehension, and that is noted and discussed in the text). That the flyer study went beyond type and layout is self-evident, but I felt readers interested in type and layout for both editorial and advertising publications would be interested in the other factors looked at.

To begin the program, I wrote two fictitious articles which were designed to have specific interest to members of the sample.

One article was about the plans for a local government authority to increase domestic rates, but to reduce the services. The article described the proposed changes in detail. This article was geographically specific to sample members and was aimed squarely at their pocketbooks.

The second article was about plans to place parking meters in suburban streets, particularly outside private houses, and to enforce on-street parking laws by constant police surveillance. It included police comments designed to inflame residents and to provoke reactions from them.

(Note that over the five years of the program, I wrote many articles about many subjects to use as test pieces; the two given here are simply examples.)

The subjects were divided into two equal groups to provide a measure of control. Each group read the same articles, but presented in alternative forms. Subjects were tested individually, in their own homes, under supervision, reading the material in a given time.

Initially, the articles were set in two design styles: one in the format of figure 10a (p.34), the other in the format of figure 10b (p.34). Half the subjects were given the Figure 10a layout, which complies with Arnold's Gutenberg Diagram. Half received the figure 10b layout, which does not.

Later, with another article, the sample groups were reversed. The first group was given the figure 10a layout, and the second group was given the figure 10b layout. This procedure was repeated several times over several years, and the sample groups were randomly selected each time from the subject pool each time. The same applied for the other reading gravity test, figure 12a and figure 12b (p.37) layouts.

For subsequent tests, serif versus sans serif, roman versus bold versus italic and so on, figure 10a became the standard layout. A few of these tests were also carried out using layout 10b too, but it was obvious that the reading gravity factor was confounding test outcomes, so it was dropped.

I conducted all the research. I devised the test pages, performed the interviews, phrased and asked the questions, and collated the results. I did it this way to eliminate the possibility of bias or distortion that might have occurred had I contracted others to do the work.

The final averages were taken from the total calculation, not as averages of a series of averages.

Assessment

The expectation was that participants who had fully read and comprehended the articles would be able to answer questions related to points throughout the them.

I had devised 10 questions about major points spread through each of the articles. When subjects had finished reading, these questions were asked in random order so as not to suggest any hierarchical relationship to the articles. My view was that if design or other factors had a detrimental effect on reading and comprehension, the point where comprehension declined would be apparent from incorrect answers.

For example, if subjects answered only two or three questions correctly, and these referred to points in the first few paragraphs of the article, it could be inferred that comprehension had failed after the first few

paragraphs. If they answer seven or more questions correctly, then clearly they had read and understood the article. When the test results showed that one group had performed poorly, on average, compared with another group, and both were reading the same material which differed only by one element or typography or design, then it could be inferred that the change in that type element or design had had the effect of lowering or improving comprehension.

Those who correctly answered from 10 to seven questions were rated as having good comprehension; six to four, fair; and three to zero, poor.

Comments

Following each formal test, I asked informal questions in an attempt to gain introspective, anecdotal evidence. The answers to these questions helped to explain some of the results. For example, those reading text set in colored type generally performed poorly; from the informal questioning, I was able to find that participants experienced eye fatigue and consequent loss of comprehension. Colored headlines distracted readers from the black text which followed; discussion afterwards revealed how the distraction occurred.

What I told participants

Insofar as the artificial test situation allowed, I wanted the subjects to approach the reading tasks in a natural manner, as they might if they were casually reading a newspaper or magazine, so I was careful not to spell out what I was trying to elucidate, nor to mislead them in a particular direction, either of which might have resulted in biased test results.

Rather, I indicated that the tests were designed to help me, as a magazine editor (the NRMA's members' magazine, *Open Road)*, to produce better material for my readers. I believe some of them assumed the improvement was to be in the content rather than the presentation, and I did not disabuse them.

My advisors

The catalyst for the methodology and procedures I used was the late Professor Henry Mayer, Professor of Political Science at Sydney University, NSW, Australia, and editor of the learned magazine *Media Information Australia*. When I mentioned to him that I was considering some research

into elements of design and typography, he gave me several invaluable pieces of advice.

Professor Mayer's first piece of advice was to produce an acceptable methodology. There would be critics, he said, mainly among those who found the results disturbed their comfortable beds of subjective prejudice. They would attack on two fronts: ad hominem, which could be countered with ease, and on the method. If the method were faulty, the attack could not be countered.

Moreover, he said, if he found the methodology faulty he would personally crucify me, irrespective of whether the results confirmed or opposed his own prejudices about design.

The next piece of advice was that I should attempt to eliminate all variables when testing a particular element. This advice led to a long sequence of experiments in which questionable design elements were progressively eliminated.

Professor Mayer also advised me to circulate the proposed methodology and results widely before publishing. If possible, he said, present them personally to a public forum, preferably to potential critics. This I did, using the Australian Creative Advertising Awards (the Caxtons), and the Australian Suburban Newspapers Association Congress, as forums.

I then sought the advice of Professor Edmund Arnold, then Professor of Mass Communications at Virginia Commonwealth University, Richmond, VA, USA, and Professor Rolf Rehe, then of the Purdue Universities Consortium in Indianapolis, IN, USA. Assisted by Professor Mayer, I adapted Rolf Rehe's rate-of-work method to suit my experiments.

Professor Mayer also advised me on sample selection.

I was advised on the conduct of the project by Dr Simon Gadir, then Director of Research for the Newspaper Advertising Bureau of Australia. Dr. Gadir also advised me on such factors as statistical significance of results, and scales of confidence (the McLemar scale) — in other words, which numbers would stand up to scientific scrutiny, and which should be (and were) eliminated.

I also sought advice on method, calculation, analysis, and presentation of results from the following:

- Professor David Sless, Executive Director of the Communications Research Institute of Australia.
- Professors Arnold, Rehe, and Mayer.
- David Ogilvy, who was a renowned researcher before he became the

doyen of modern advertising.

- Bryce Courtenay, who when he is not writing bestselling novels and screenplays (*The Power of One*), was Creative Director for the George Patterson Advertising Agency in Sydney.
- Members of the academic staff of the University of Reading, UK, which houses the British Government's forms design center.
- Members of the academic staff of the Royal College of Arts in London.
- Jim Thomson, then Head of Research for the National Roads and Motorists' Association (NRMA), Australia, one of the ten biggest motoring organizations in the world.

Comprehension or readability?

Over the years of my research, David Ogilvy more than once raised the question whether I was measuring reading comprehension or merely readability. I'm grateful for his persistence in moving me to confront this important and difficult question.

My difficulty stemmed from the fact that, in my Shorter Oxford Dictionary, readability has two distinct definitions:

1. Capable of being read, legible.
2. Capable of being read with pleasure or interest, usually of a literary work.

This presented me with a dilemma. I was not testing legibility; that is a totally different discipline and has been the subject of a considerable measure of scientific work. I wanted to know whether the nature of type or design affected the reader's ability to *understand* the text.

However, if I used the words readability and readable, they could be misconstrued as being analogous to legibility and legible.

So I settled on comprehensibility.

The essence is, as David Ogilvy suggests, that I was measuring the extent to which typography and design affected understanding. But to understand, one first has to read the message

However, in Chapter 6, where I address the subject of headlines, I did not test comprehension, because the method did not permit it. There, I was testing whether headlines were easy to read or not. Those were the words used in the tests, and those were the words used by the sample participants.

COLOR, READING, EYES Appendix 2

On color, type, reading and the eyes

By Geoffrey Heard

The test outcomes which show that almost any departure from the standard of black serif type on white (or light) paper results in some interference to the reader's ability to read and comprehend text, are consistent with the results of psychological experimentation in these and related fields.

Certainly, when I first read Colin Wheildon's book, it gave rise to a number of "Aha!" moments in my mind. My studies in psychology included vision, information acquisition, reading and information processing in the brain. "Aha!", I cried, the moment I saw Colin Wheildon's results on the use of high chroma colors in headlines, then "Aha!" I cried again when I read the results of reading non-black/dark body type and color on color, and yet again "Aha!" when I read the outcomes of his tests on type set justified, ragged left and ragged right.

Color

The fact that high chroma colored headlines keep intruding into reading of the body type under them fitted perfectly with the notion of the primacy of color, hard-wired into very low levels of the human brain, over the learned skill of reading. Psychologists have speculated that just such colors have been the signs of danger for humankind reaching far back down the evolutionary chain, so when we see them they absolutely demand attention. Certainly, recognition of such colors is wired into very low levels of the human brain — and for good reason. Think of the nocturnal animal from which we evolved, scuttling around in the jungle. In that environment, red and other high chroma colors often are the signs of danger — toxic plants and venomous animals. Back in World War II, Australian and American personnel on active service in the Southwest Pacific theatre were issued with a little booklet to help them live off the land if they were forced to do so. Called *Friendly Fruits & Vegetables*, it carried this dire warning on page 3: "In general, red means danger in the jungle, so