04 The Influence of Serifs on 'h' and 'i': Useful Knowledge from Design-led Scientific Research

DR. SOFIE BEIER, DR. MARY C DYSON

ABSTRACT

The typographical naivety of much scientific legibility research has caused designers to question the value of the research and the results. Examining the reasons underlying this questioning, the paper discusses the importance of designers being more accepting of scientific findings, and why legibility investigations have value. To demonstrate how typographic knowledge can be incorporated into the design of studies to increase their validity, the paper reports on a new investigation into the role of serifs when viewed at a distance. The experiment looks into the identification of the lowercase letters 'j', 'i', 'l', 'b', 'h', 'n', 'u', and 'a' in isolation. All of the letters originate in the same typeface and are presented in one version with serifs and one version without serifs. Although the experiment found no overall legibility difference between the sans serif and the serif versions, the study showed that letters with serifs placed on the vertical extremes were more legible at a distance than the same letters in a sans serif. These findings can therefore provide specific guidance on the design of individual letters and demonstrate the product of collaboration between designer and scientist on the planning, implementation, and analysis of the study.

In recent years, a growing interest in strengthening the connection between the two different ways of working is emerging, with various research institutions encouraging interdisciplinary work.

When the field of typeface legibility research functions at its best, it contains elements of both science and design.

The scientific approach contributes with controlled test methods and with the analysis of the data, while the design approach contributes with the creation of relevant material for testing.

THE PROBLEMS OF THE TYPEFACE LEGIBILITY FIELD

A number of experimental legibility studies comparing different typefaces were carried out in the first part of the 20th century (for English-language ones see: Tinker 1964; Luckiesh & Moss 1942; Pyke 1926; Roethlein 1912). In recent times, a large number of researchers have also published papers related to legibility (for examples see: Fiset et al. 2008; Sheedy et al. 2008; Fox et al. 2007). These studies are published in academic journals, which are often not read by the practising designer. Even if they were read, the way the material is presented often makes it difficult for practitioners to understand the findings and translate them into usable knowledge. In 1969, Merald Wrolstad, the editor of the *Journal of Typographic Research* (now *Visible Language*), explained the lack of interest in research into legibility among designers. She pointed out that since research projects tend to be motivated by the researcher's area of interest, and because many legibility researchers are engineers, psychologists, or reading specialists, the research will focus on variables of interest to these disciplines and not on the various aspects of letter-forms.

An argument voiced by a number of designers through the online typography community Typophile (2008a; 2008b) is that we must fully understand the underlying principle of legibility before studying it. This argument fails to recognize that understanding is acquired through research, and building on existing knowledge. The claim made in these threads is that no test method that includes typefaces as test material will be able to provide any useful data, due to the difficulties in isolating a single typographic stylistic feature for investigation.

Hence they argue that researchers do not understand the principle of legibility because they use typefaces in their studies that vary in many ways.

To make matters more complicated, it is fairly common that the type size on the body varies considerably between typefaces – a variation that has existed since the early printers. In such cases, one typeface can have a letter height of 8 millimetres when set in 32 points, while another typeface in the same point size will have a letter height of 6 millimetres (*figure 2*). Equating point size is consequently not an appropriate means of comparing different typefaces.

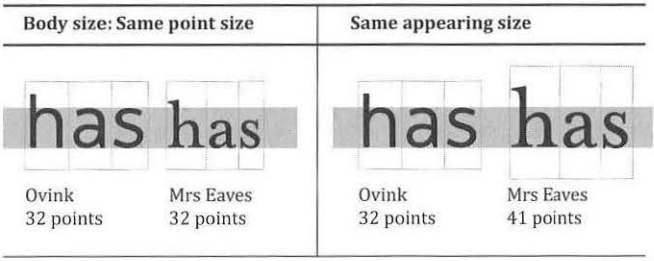

FIGURE 2 *The typeface Mrs Eaves is set at an unusually small size on the body. In this case the problem of equivalence is not only related to the x-height but to the extending parts of the letter as well. To the left are the typefaces Ovink and Mrs Eaves both set in 32 points; to the right the same typefaces are adjusted to have a similar appearing size. This results in different point sizes, and ascenders and descenders having different lengths.*

A second potential problem with test material that compares different typefaces is that typefaces tend to vary in both weight and width. There is a risk of drawing incorrect conclusions if the researcher is not aware of the possible implications of these variables. The fact that they do influence legibility is demonstrated in an experimental investigation by Barbara Elizabeth Roethlein (1912). Looking at the relative distance legibility of a number of different styles of the typeface Cheltenham, the study found that both typeface weight and width influence performance when text is read at a distance. This is an early example of a controlled investigation where the differences in stylistic features are kept to a minimum.

The raw data presented in the paper shows that the Bold style could be read at a greater distance than both the Regular (Ordinary) and the Bold Condensed style. When comparing the data for the Regular and Bold styles, we can be fairly sure that the difference in performance originates in the difference in weight. Similarly, we can be fairly sure that the difference in performance between the Bold and the Bold Condensed styles originates in the difference in width.

The results of Roethlein's study mean that if two different typefaces were compared at a distance and these happened to vary in weight and width, as well as other stylistic features, a style with heavier

serifs on ascenders and descenders may help ease reading in the parafoveal view by emphasising these specific parts of the letters. Type designer Adrian Frutiger (n.d.) has a similar argument, suggesting that a stroke with no fortified endings will leave the observer with the feeling that it could flow on forever. He finds that serifs, like the base and the top of architectural columns, give emphasis to the terminations of the column. Another central matter often referred to when speaking in favour of the serifs, is the role they play in facilitating a higher differentiation between letters. According to designer Erik Spiekermann (2007), serifs helps to avoid confusions between the lowercase letters 'l', 'i', the uppercase 'I', and the digit '1' when these letters are in isolation.

However, the majority of reading rate experiments that have compared the legibility of sans serif and serif typefaces have found the performance of the two to be similar. In a critical review of 72 studies that compare different typefaces, Ole Lund (1999) found no valid conclusion in favour of either serifs or sans serif typefaces. However, an experiment carried out by mathematician Robert A. Morris with vision scientist colleagues (2002) has shown that a serif version of the typeface Lucida, of 40 pixels x-height presented at a 4 meter distance, slows down reading compared to a version of Lucida without serifs and with slightly less stroke contrast. The difference between this study and the many it succeeded is that one of Morris's co-authors is a type designer (Charles Bigelow).

By applying an interdisciplinary approach between fields of science and design, the research produces more rigorous and relevant findings than when the work is informed by one discipline alone.

The findings by Morris and colleagues indicate that serifs are not always either good or bad. An investigation by John Harris (1973) may explain this ambiguity, as his results suggest that serifs seem to reduce the legibility of certain letters, and improve others. In a study of the sans serif typefaces Univers 689 and Gill Sans Medium, and of the serif typeface Baskerville 169, Harris applied the methodology of briefly presenting a single letter four degrees off the centre of vision, and found that the confusions of 'b' > 'h', 'a' > 'n', 'n' > 'u', and 'u' > 'n' were significantly higher in the serif typeface compared to the two sans serif typefaces (*figure 3*).

compared to the performance of a sans serif 'i' of the same typeface. However when the serifs are placed at both the top and the baseline, the identification of the character is just as poor as the version with no serifs at all (*figure 5*).

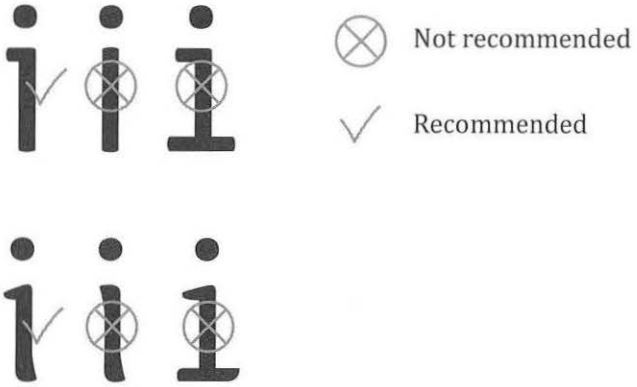

FIGURE 5 *Two of the typefaces tested by Beier & Larson (2010) (top row Ovink, lower row Spencer). Three versions of the letter 'i' were tested in each of the typefaces*

The study by Morris and his colleagues (2002) compared the letters as a whole without looking at individual letters; Harris (1973) looked at individual letters, yet applied the problematic method of comparing different typefaces; Beier & Larson (2010) compared different design versions of letters within the same typefaces, yet only examined the effect of serifs on the letter 'i'.

This study looks into the legibility of the eight letters 'j', 'i', 'l', 'b', 'h', 'n', 'u', and 'a' identified by Harris (1973) as being influenced by serifs in either a positive or negative direction. The investigation looks at the confusions between these letters both with and without serifs. Following the analysis by Ole Lund (1999) who found no valid conclusion in favour of either serifs or sans serif typefaces, no overall legibility difference between the serif and the sans serif typeface styles is expected. Instead, we wish to investigate whether serifs at the vertical extremes have a positive influence on legibility (Unger 2007; Frutiger n.d.). Therefore the letters l', 'b', 'h', 'n', 'u', all having serifs at their vertical extremes, are compared in their serif and sans serif styles (*figure 6*). It is further expected that the results of the letter confusion comparison will show that some placements of serifs facilitate high legibility and others do not. It is interesting to determine whether the findings are similar to those reported by Harris (1973), or whether a better control of the variables produces a different result.

specific reading situation that is under investigation. The focus of the present investigation is the role serifs play when viewed from straight ahead at a distance.

METHOD

In this study, the letters used as test stimuli are designed by the first author to control the variation between the objects of study (the two typeface styles). A sans serif typeface designed for high distance legibility is based on the regular weight of the typeface Ovink (from here on referred to as OvinkSans). For a serif comparison, a new slab serif typeface style named OvnkSerif was designed. The single difference between the letters of the two typeface styles is the added serifs to the typeface OvinkSerif. All other aspects are identical across the two styles (*figure 7*).

FIGURE 7 *The two typeface styles of OvinkSans and OvinkSerif superimposed*

The reason for creating a slab serif, rather than a traditional serif typeface, relates to stroke contrast and demonstrates the manner in which design knowledge informs the choice of suitable material for testing. Traditional serifs have a higher stroke contrast than sans serif typefaces. If the serif typeface were of a traditional nature instead of slab serif, both the sans serif and the serif style would have needed a larger stroke contrast than that of OvinkSans. Large stroke contrast in sans serif typefaces is unusual and would therefore have resulted in unnatural test material. A slab serif type was chosen because it can have the same stroke contrast as sans serif type.

It is likely that the performance difference between two conditions that have only one variable that differs is small compared with conditions of test material that have several stylistic features that vary simultaneously (as in much previous research). In the isolation of properties, there is a risk that the difference between test materials is too small for a given method to detect. Yet, this should never be an argument

English alphabet would be shown with different numbers of occurrences. In order to minimize eyestrain, participants were asked to take as many breaks as they felt necessary. Each letter was presented on white paper at a point size of 22 (with an x-height of 4 mm). The ambient room light was typical for an office environment.

RESULTS

The average number of correct identifications of the letters 'j', 'i', 'l', 'b', 'h', 'n', 'u', and 'a', across participants is shown in Table 1 for each typeface. This data illustrates which letters are more accurately identified across styles but does not reveal confusions or take account of systematic response biases. For example, if a participant has a tendency to respond 'i', they are likely to get a high correct identification of the letter 'i'. However, their false alarms (i.e. saying 'i' when presented with a different letter such as 'l') have not been taken into account.

	j	i	l	b	h	n	u	a
OVINKSANS	2.31 (0.49)	3.92 (0.40)	1.00 (0.36)	2.38 (0.46)	2.46 (0.35)	2.92 (0.33)	2.08 (0.45)	1.54 (0.42)
OVINKSERIF	2.15 (0.36)	3.08 (0.43)	2.08 (0.45)	2.62 (0.40)	3.00 (0.45)	3.69 (0.31)	2.46 (0.51)	1.62 (0.43)

TABLE 1 *The means (out of a possible 5) of the correct identifications of the 8 letters under investigation, in each typeface style. Standard errors of the means are in parentheses.*

By calculating p(A) using hit and false alarm rates, a measure of sensitivity is obtained which is free of response bias (McNicol, 1972, 113). A value of 0.5 indicates chance performance and 1 equals perfect performance. Table 2 shows the values of p(A) for each letter in each typeface style with the false alarm based on the number of times that any of the other 25 letters is identified as that letter. This represents a more sophisticated analysis of the data than number correct, providing a more accurate account of letter recognition.

Analysis of the data in Table 2 was carried out correcting degrees of freedom for sphericity using the Greenhouse-Geisser estimate of sphericity ($\Sigma = 0.52$). This shows an effect of letter ($F(7, 84) = 4.17$, $p = 0.0073$), as would be predicted from Tinker (1964). There is no overall difference between the OvinkSans and OvinkSerif ($F(1, 12) = 0.88$, $p = 0.36$).

incorrect responses with serifs, but some exceptions. The pairs l > i and i > l and the pairs b > h and h > b appear to account for the interaction. In particular, 'l' is misreported as 'i' more often in OvinkSans, whereas 'i' is misreported as 'l' more often with OvinkSerif. 'b' is misreported as 'h' a similar number of times in both typeface styles, whereas 'h' is misreported as 'b' more often in OvinkSerif.

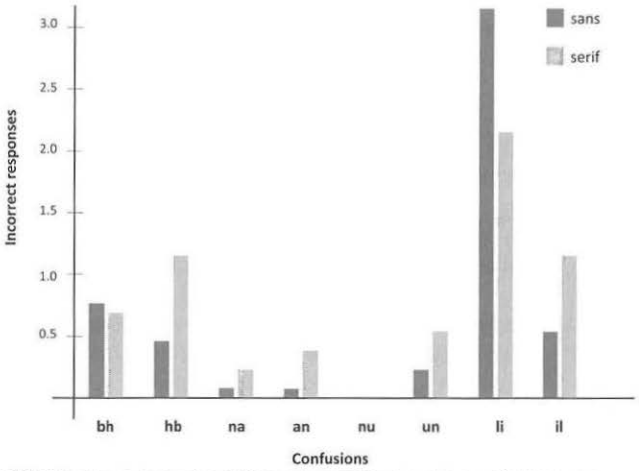

FIGURE 9 *The mean number of incorrect responses between specific letter pairs.*

DISCUSSION

As expected, the overall data for the two typefaces show no statistical difference between OvinkSerif and OvinkSans, indicating that when functioning as whole alphabets, sans serif and serifs typeface styles are equally legible. However, this result obscures the different effect that serifs may have on individual letters, revealed by Harris (1973); when looking at the letters in groups, there are some more subtle findings. The theory put forward by Unger (2007) and Frutiger (n.d.) that serifs at the extremes help facilitate letter recognition, is confirmed in the collective data of the group of letters with serifs at their vertical extremes ('l', 'b', 'h', 'n', and 'u'). The data shows the serif letters were identified significantly better than the sans serif letters.

This indicates that serifs at the vertical extremes do, in fact, facilitate a better definition of the stroke and through this enhance the recognition of letter features.

We believe that an investigation such as this is far more useful for designers than investigations that, for example, compare the speed of reading Helvetica to the speed of reading Times. The result of such a comparison is only valid to the designer choosing between Helvetica and Times, and is irrelevant in any other situation. It does not enlighten us as to the legibility of the individual features of the typeface style and why participants read the letters the way they do, or to help define the features that can improve the speed of reading in general. Studies that follow the principle of isolating typographical variables contribute more to the field of design. The research not only informs us as to whether one typeface style is more legible than another typeface, but presents findings that help identify the role that different letter features play in a given context.

CONCLUSION

The present investigation found no difference in the distance legibility of sans and serif typeface styles when studying the collective results of the individual letters 'j', 'i', 'l', 'b', 'h', 'n', 'u', and 'a'. However, looking at a group of letters with serifs at the vertical extremes, the experiment demonstrated higher distance legibility when these letters have serifs on the stems; this demonstrates that serifs on such locations play a central role in facilitating high distance legibility. The data further indicates that serifs on the counter of 'h' and on both ends of the stem of 'i' cause a higher misreading for 'b' and 'l' respectively. The findings as a whole suggest that serifs should not always be applied in the conventional fashion as seen in traditional Old Style and Didone typefaces; instead they have an ability to facilitate higher distance legibility if they are placed in a semi-serif fashion where they are positioned at relevant stroke endings.

The results were reached through an interdisciplinary collaboration between researchers from the fields of science and design. By designing typefaces that allow for isolation of a single typographical variable, and through the creation of an experiment that adopts a scientific approach, this study achieves a high level of internal validity. By adopting a research question that is relevant to the design community, the research further provides a high level of external validity. The detailed and specific nature of these results can inform character design by identifying features that influence letter recognition. This paper has aimed to provide a foundation for future research through espousing the benefit of integrating scientific and design expertise,and by providing an example of the application of this approach. Further work could build upon these findings to pursue the role of serifs in more detail or to investigate other typographic stylistic features of typefaces (e.g. weight or stroke contrast) in relation to particular reading situations.

Fox, Doug., Chaparro, Barbara S., & Merkle, Ed. 2007. Examining Legibility of the Letter "e" and Number "0" Using Classification tree Analysis. *Usability News,* 9(2), http://www.surl.org/usabilitynews/92/legibility.asp (Accessed February 1, 2013).

Frutiger, Adrian (n.d.), The History of Linear, Sans Serif Typefaces, http://www.linotype.com/2258-16893/thelossoftheserifspart1.html (Accessed February 15, 2013).

Harris, John. 1973. Confusions in Letter Recognition. *Professional Printer.* 17.2, 29-34.

Hess, R.F., Dakin, S.C. & Kapoor, N. 2000. The Foveal 'Crowding' Effect: Physics or Physiology?. *Vision Research.* 40, 365-370.

Heylighen, Ann., Cavallin, Humberto., & Bianchin, Matteo. 2009. Design in Mind. *Design Issues.* 25.1, 94-105.

Larson, Kevin & Sheedy, James E. 2008. Blink: The Stress of Reading. Eye. 67. 95.

Liu, Lei. & Arditi, Aries. 2001. How crowding affects letter confusion. *Optometry and Vision Science.* 78.1, 50-55.

Lund, Ole. 1999. Knowledge Construction in Typography: The case of legibility research and the legibility of sans serif typefaces. Thesis PhD, The University of Reading, Department of Typography & Graphic Communication.

Luckiesh, Matthew. & Moss, Frank K. 1942. *Reading as a Visual Task,* New York: D. Van Nostrand Company.

Mansfield, J. S., Legge, G. E., & Bane, M. C. 1996. Psychophysics of reading: XV. Font effects in normal and low vision. *Investigative Ophthalmology & Visual Science,* 37, 1492-1501.

McClelland, James L. & Rumelhart, David E. 1981. An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings. *Psychological Review.* 88.5, 375-407.

McNicol, D. 1972 *A Primer of Signal Detection Theory.* London: George Allen & Unwin Ltd.

Morris, Robert A., Aquilante, Kathy., Yager, Dean. & Bigelow, Charles. 2002. P-13: Serifs Slow RSVP Reading at Very Small Sizes, but Don't Matter at Larger Size. *SID Symposium Digest of Technical Papers.* 33.1, 244-247.

ABOUT THE AUTHORS

Dr. Sofie Beier is a type designer and assistant professor employed by the School of Design under The Royal Danish Academy of Fine Arts. She holds a PhD from the Royal College of Art in London, and is the author of the book *Reading Letters: Designing for Legibility*. Her research focuses on improving the reading experience by gaining a better understanding of how different typefaces and letter shapes can influence the way we read.

www.sofiebeier.dk

Dr. Mary Dyson is Senior Lecturer in the Dept of Typography & Graphic Communication, University of Reading, UK. She studied experimental psychology leading to a PhD in perception, and now teaches theoretical and empirical approaches to typography and graphic communication. She has supervised many research students over the years on quite diverse topics and continues to enjoy this activity. Her most recent research interests are driven by a desire to bridge the gap between scientists and designers and find commonalities. Her experiments have looked at issues such as how designers perceive typefaces compared to non-designers and how Chinese typefaces are perceived by non-Chinese readers, with the aim of understanding how typefaces are processed in reading.

http://www.reading.ac.uk/typography/about/staff/m-c-dyson.aspx

AUTHOR CONTACT

Dr. Sofie Beier
School of Design
The Royal Danish Academy of Fine Arts
Philip de Langes Allé 10
DK-1435 København K
Denmark
sbe@kadk.dk

Dr. Mary C Dyson
Department of Typography & Graphic Communication
University of Reading
2 Earley Gate
Whiteknights
Reading RG6 6AU
M.C.Dyson@reading.ac.uk