Legibility: how and why typography affects ease of reading

In these pages you will find an explanation of how we reads the movement of our eyes when we read, and how we recognize words. You will find material of a diverse nature and complexity that will help explain the basic principles of typography, wrong ideas and practices that are common in the profession, and advanced concepts of reading on screen supported by scientific research.

Table of contents

Chapter 1

What do we mean by legibility?

Broad definition

When I mentioned to a couple of people that I intended to write a text on legibility they asked ‘legibility of what?’ The answer is legibility of text, but the question may have been looking for a more specific focus, i.e. what type of texts. The question also encouraged me to reflect on a more general interpretation of legibility. For example, the phrase from a dictionary which illustrates another use: ‘an anxious mood that was clearly legible upon her face’ shows that we read people’s facial expressions and interpret their mood from these. Although it may be intriguing to read faces, I intend to focus on reading text, and in particular ease of reading.

Within typographic and graphic design, we might consider whether signs are legible (in particular from a distance), whether we can decipher small print (especially later in life), if icons can be easily identified or recognised (without text labels), if a novel or textbook is set in a readable type (encouraging us to read on). These questions emphasise that it is not only the physical characteristics of the text or symbol that need to be considered in determining whether or not the designs are legible, or how legible they are. The purpose for reading, the context of reading, and the characteristics of the reader also determine legibility.

Question: Is legibility a binary concept (i.e. legible or illegible) or are there degrees of legibility, and perhaps also illegibility? If there are degrees, how do we decide what is an acceptable level of legibility? (We will return to this question in the final chapter.)

In describing various examples of designed objects, I have used adjectives other than 'legible' to describe the ease of reading, e.g. being identifiable, recognisable, or readable. These terms may be helpful in conveying the general meaning of legibility but there are circumstances where it is important to differentiate among them, and to be more precise in our definition. For example, when evaluating research, it is necessary to know what operational definition of legibility has been used by the researchers. An operational definition describes what is measured in the study (see Chapter 4).

Legibility, readability, and related concepts

Another way of considering what is meant by legibility is to distinguish it from related concepts. Starting with the initial sensation of an image on our retina, part of our eye (see Figure 2.2, Chapter 2), for this image to register, it must be 'visible' or 'perceptible' [1]. If it is too far away, for example, it will not be perceptible. We may therefore consider visibility or perceptibility as a prerequisite for legibility: if something is not visible, it cannot be legible. It may not always be possible to make a clear distinction between where perceptibility stops and legibility begins and this will become clearer when reviewing the methods used to test legibility (Chapter 4: Threshold and related measures. I will therefore include perceptibility as part of legibility.

Another distinction can be made between legibility and readability. Some authors, notably typographer, writer, and designer Walter Tracy, make the point that legibility and readability of type are separate attributes: legibility refers to the clarity of individual characters; readability refers to the ease with which we comprehend a text (Tracy 1986, p31). Unfortunately this definition of readability can be rather confusing as comprehension is influenced by typographic form, but also the complexity of the content affects our understanding of a text. For this reason, I am going to use a single concept, 'legibility', which will cover:

identifying individual characters, whole words, and reading text which will usually refer to continuous texts for extended reading, typically sentences arranged into paragraphs and sections.

This book is a good example of continuous text, although it is interrupted by other text elements (e.g. lists) and illustrations. I think it is too ambitious to expand the scope of this book to non-continuous texts, such as tables, signs, and forms, but I will mention signs from time-to-time.

I consider it important to update our definition of legibility to take into account that we frequently read from screens. Text can be read in print or on screen and usability may be a better way of describing the ease of working with print or screen documents, which may be affected by the layout or interface design. The term usability typically incorporates navigation and other forms of interaction with the text, as well as reading. Although I am focusing primarily on reading text and legibility, there may be some overlap with usability. The important point is to clarify what is measured in a study, rather than the particular word used by the researchers as these may differ.

?

Question: Which design variables might influence the legibility of this book?

By offering a fairly loose, and rather general, description of legibility, I wish to avoid getting too involved in analysing differences among definitions. Instead, we might consider how definitions highlight various characteristics or criteria and contribute to a fuller description which encompasses how legibility is measured and the context of reading. Panel 1.1 introduces several definitions from different sources, and I will return to some of these in later chapters.

Panel 1.1: Definitions of legibility

The Concise Oxford dictionary lists two criteria:

- 'clear, capable of being read'

The Shorter Oxford English Dictionary:

- Of writing: plain, easily made out
- Of compositions: accessible to readers; also easy to read, readable

Reynolds and Simmonds (1984, p1) provide a fuller definition which refers to the nature of the material and differentiates between ease and speed:

- 'ease and speed of recognition of individual letters or numerals, and of words either singly or in the form of continuous text'

Zachrisson (1965, p25) incorporates comprehension and includes a measure of accuracy:

- 'the speed and accuracy of visually receiving and comprehending meaningful running text'

Williamson (1983, p378) uses similar measures, but described in slightly different ways, and also introduces an environmental factor referring to the circumstances of reading:

- ‘the ability to read comfortably, continuously and swiftly by intended reader in appropriate circumstances’

Finally, a very different way of considering legibility is offered by Gill (1931, p47):

- ‘legibility, in practice, amounts simply to what one is accustomed to’

And subsequently Licko (1990, p13) expressing the same sentiment:

- ‘You read best what you read most’

Evidence for legibility

It is therefore important to be critical of evidence that supports particular positions. We should question what the evidence is and how it was obtained. In the exercise above, you used your experience to make judgements about legibility. These judgements are useful and sometimes form part of legibility research.

An issue for discussion is whether designers can make claims concerning legibility if they have no means of supporting their claim other than their own judgement. I do not underestimate the value of professional knowledge, craft experience, or practical design skills and training. However, at the very least, I believe it is important to check that we have not developed less than optimal ways of presenting text which may be based on misguided notions of what readers find easiest to read.

Question: In your opinion, what contribution can designers' judgements make to determining what is most legible?

In this text I am going to focus on empirical research, commonly studies testing different typographical arrangements on a group of participants. Most of the research is based on adult reading but occasionally I describe some studies which include children because the typography may need to be different to cater for the developing reader.

Why is legibility important?

Legibility focuses the designer on the functional characteristics of a text to make a message accessible. There has been some opposition to legibility research, or even prioritising functionality, but this tends to be criticism of the methods used, and consequently what is measured (discussed further in Chapter 3 and 4). When the purpose is to convey a message, one of the roles of typography is to support reading.

Legibility is one aspect of universal or inclusive design, which is designing to meet the needs of people of diverse age and capability. In the UK, the British Standards Institute introduced a standard in 2005 and defined inclusive design as:

> *The design of mainstream products and/or services that is accessible to, and usable by, people with the widest range of abilities within the widest range of situations without the need for special adaptation or design.*
>
> British Standards Institute (2005, p4)

By designing legible material, we are supporting the ability of people to complete activities and tasks. The Web Accessibility Initiative explains the close relationship between accessibility, usability, and inclusive design.

Functionality versus aesthetics

A classic lecture given by Beatrice Warde in 1930 presents the case for ‘invisible type’, meaning the reader should not notice the characteristics of the type (Warde, 1930 in Armstrong, 2009, p41) as these may detract from communicating the message [2] . This ideal appears to be in opposition to aesthetic considerations, if we interpret aesthetics as the creation of a beautiful text which draws attention to the typography. However, an alternative proposition is that legible text is also aesthetically pleasing. Therefore, legibility and aesthetics need not be seen as opposing aims in the design of continuous text.

> *...all designing — whether a car, a coffee pot, or a typeface — is a process in which two aspects should combine and balance: the object must work well, and it must look well.*
>
> Tracy replying to Donald E. Knuth's article “The concept of a meta-font”, (1982, p355)

Another view of the relationship between functionality and aesthetics is that typefaces have both these roles: a functional role relating to legibility; and an aesthetic or semantic role which determines whether the typeface is suitable for certain purposes because of the meaning conveyed by the visual form. This second role has been described using different terms: atmosphere value, congeniality, semantic qualities, and personality. More recently research has extended beyond typefaces to look at impressions gained from different typographic layouts, and 'interaction aesthetics' which are emotions emerging from interacting with products.

Question: Are there some objects or systems created by graphic designers where aesthetic considerations may be more important than legibility?

Exercise: rank the three examples (Figures 1.1A, 1.1B, 1.1C) according to your judgement of their legibility with 1 the most legible and 3 the least legible. Now rank the same three examples according to your aesthetic judgement with 1 the most pleasing.

Ibussis achucte molutera, inpracchuit videffrei ine mendam pri inatatq uodis. Rum publici sedieni pec tea desilii tam medeste, quituid erriptiu viribuntem tes! Is. Ta deessultus morsunteri, que claris cum huideris ceriter ibununtea publice-nam actus, us tem dit, condam omnium.

Inte essit. Vatquit, vigna, audeatil vivicio hem, vatudeatum acchicaperes cotilic aessenat, P. Sate in Ita ia rei tem pul ut publicae, quius, tatus? Hucit L. M. Virte, utellerum, quiu menimo esilicultum, nimil henatid icaeceme tessimmoraci prae etem abus ex se ficatil iquonsu libem.

Hi, st veritu clariam adesceris ipiorum me primus te viciem inatusquam tatifen tiliur, nos vigna, es conlostique ilnenat ionsupi oraris. Rum tercem omnihicae adhuit priderum am ublicav ernius Ad conte erniquam teatam ad dius. Upiem res ad consilia esuam dis cam.

Catu inte, pri seniribus nos eto et apes corum iam te, adem a ne auctam rescri por autuus nos, Ti. Essus ste culatua nem nonsulist det? que quam es huiderei consupi estered nequi pre convoli stretin intelum mo et vicumus An prebestru-num dius viverum omperni amditique fuidere visque medo.

Figure 1.1A

Ibussis achucte molutera, inpracchuit videffrei ine mendam pri inatatq uodis. Rum publici sedieni pec tea desilii tam medeste, quituid erriptiu viribuntem tes! Is. Ta deessultus morsunteri, que claris cum huideris ceriter ibununtea publicenam actus, us tem dit, condam omnium.

Inte essit. Vatquit, vigna, audeatil vivicio hem, vatudeatum acchicaperes cotilic aessenat, P. Sate in Ita ia rei tem pul ut publicae, quius, tatus? Hucit L. M. Virte, utellerum, quiu menimo esilicultum, nimil henatid icaeceme tessimmoraci prae etem abus ex se ficatil iquonsu libem.

Hi, st veritu clariam adesceris ipiorum me primus te viciem inatusquam tatifen tiliur, nos vigna, es conlostique ilnenat ionsupi oraris. Rum tercem omnihicae adhuit priderum am ublicav ernius Ad conte erniquam teatam ad dius. Upiem res ad consilia esuam dis cam.

Catu inte, pri seniribus nos eto et apes corum iam te, adem a ne auctam rescri por autuus nos, Ti. Essus ste culatua nem nonsulist det? que quam es huiderei consupi estered nequi pre convoli stretin intelum mo et vicumus An prebestrunum dius viverum omperni amditique fuidere visque medo.

Figure 1.1B

I bussis achucte mo lutera , i n p racchuit videffrei ine mendam pri inatatq uodis. Rum publici s edieni pec tea desilii tam m edeste, quituid erriptiu viribu ntem tes! Is. Ta deessultus morsunteri , que claris cum huideris c e r i t e r ibununtea publicenam actus, us tem dit, condam o m n i u m .
Inte essit. Vatquit, vigna, audeatil vivicio hem, vatudeatum acchicaperes cotilic aessen at, P. Sate in Ita ia rei tem pul ut publicae, q u i u s , tatus? Hucit L. M. Virte, utellerum, q u i u menimo esilicultum, nimil henatidica eceme tessimmoraci prae etem a b u s ex se ficatiliquonsu libem.
Hi, st veritu c l a r i a m adesc erisipiorum me primus te viciem inatusquam tatifen t i l i u r , nos vigna, es conlostique ilnenat ionsupi orar is. Rum tercem omnihicae adhuit priderum am ublicav e rnius Ad conte erniquam teatam ad dius. Upiem res ad consilia esuam dis cam.
Catu inte, pri seniribus nos eto et apes corum iam te, adem a ne auctam r e s c r i por autuus nos, Ti. E s s us ste culatua nem nonsulist det? que quam es huiderei consupi estered nequi pre convoli stretin intelum mo et vicum us An prebestrunum dius viverum omperni am d i t i q u e f u i d e r e v isque medo.

Figure 1.1C

You may find that your judgements of legibility coincide with your judgements of what is the most aesthetically pleasing.

Now do the same set of two rankings for these 3: Figures 1.2A, 1.2B, 1.2C.

Ibussis achucte molutera, inpracchuit videffrei ine mendam pri inatatq uodis. Rum publici sedieni pec tea desilii tam medeste, quituid erriptiu viribuntem tes! Is. Ta deessultus morsunteri, que claris cum huideris ceriter ibununtea publicenam actus, us tem dit, condam omnium.

Inte essit. Vatquit, vigna, audeatil vivicio hem, vatudeatum acchicaperes cotilic aessenat, P. Sate in Ita ia rei tem pul ut publicae, quius, tatus? Hucit L. M. Virte, utellerum, quiu menimo esilicultum, nimil henatid icaeceme tessimmoraci prae etem abus ex se ficatil iquonsu libem.

Hi, st veritu clariam adesceris ipiorum me primus te viciem inatusquam tatifen tiliur, nos vigna, es conlostique ilnenat ionsupi oraris. Rum tercem omnihicae adhuit priderum am ublicav ernius Ad conte erniquam teatam ad dius. Upiem res ad consilia esuam dis cam.

Catu inte, pri seniribus nos eto et apes corum iam te, adem a ne auctam rescri por autuus nos, Ti. Essus ste culatua nem nonsulist det? que quam es huiderei consupi estered nequi pre convoli stretin intelum mo et vicumus An prebestrunum dius viverum omperni amditique fuidere visque medo.

Figure 1.2A

Ibussis achucte molutera, inpracchuit videffrei ine mendam pri inatatq uodis. Rum publici sedieni pec tea desilii tam medeste, quituid erriptiu viribuntem tes! Is. Ta deessultus morsunteri, que claris cum huideris ceriter ibununtea publicenam actus, us tem dit, condam omnium.

Inte essit. Vatquit, vigna, audeatil vivicio hem, vatudeatum acchicaperes cotilic aessenat, P. Sate in Ita ia rei tem pul ut publicae, quius, tatus? Hucit L. M. Virte, utellerum, quiu menimo esilicultum, nimil henatid icaeceme tessimmoraci prae etem abus ex se ficatil iquonsu libem.

Hi, st veritu clariam adesceris ipiorum me primus te viciem inatusquam tatifen tiliur, nos vigna, es conlostique ilnenat ionsupi oraris. Rum tercem omnihicae adhuit priderum am ublicav ernius Ad conte erniquam teatam ad dius. Upiem res ad consilia esuam dis cam.

Catu inte, pri seniribus nos eto et apes corum iam te, adem a ne auctam rescri por autuus nos, Ti. Essus ste culatua nem nonsulist det? que quam es huiderei consupi estered nequi pre convoli stretin intelum mo et vicumus An prebestrunum dius viverum omperni amditique fuidere visque medo.

Figure 1.2B

Ibussis achucte molutera, inpracchuit videffrei ine mendam pri inatatq uodis. Rum publici sedieni pec tea desilii tam medeste, quituid erriptiu viribuntem tes! Is. Ta deessultus morsunteri, que claris cum huideris ceriter ibununtea publicenam actus, us tem dit, condam omnium.

Inte essit. Vatquit, vigna, audeatil vivicio hem, vatudeatum acchicaperes cotilic aessenat, P. Sate in Ita ia rei tem pul ut publicae, quius, tatus? Hucit L. M. Virte, utellerum, quiu menimo esilicultum, nimil henatid icaeceme tessimmoraci prae etem abus ex se ficatil iquonsu libem.

Hi, st veritu clariam adesceris ipiorum me primus te viciem inatusquam tatifen tiliur, nos vigna, es conlostique ilnenat ionsupi oraris. Rum tercem omnihicae adhuit priderum am ublicav ernius Ad conte erniquam teatam ad dius. Upiem res ad consilia esuam dis cam.

Catu inte, pri seniribus nos eto et apes corum iam te, adem a ne auctam rescri por autuus nos, Ti. Essus ste culatua nem nonsulist det? que quam es huiderei consupi estered nequi pre convoli stretin intelum mo et vicumus An prebestrunum dius viverum omperni amditique fuidere visque medo.

Figure 1.2C

I have included this second set of examples to demonstrate that legibility and aesthetics may not always coincide. This may seem to contradict my proposition above, but I include it to illustrate that demonstrations can be quite convincing, until a counter example is provided that is equally convincing.

Summary

When applied to reading, legibility has been described in many ways and there are disagreements about:

- whether or not it should apply only to individual characters
- how it is distinguished from readability (or other related terms)
- its relationship with aesthetics
- how relevant legibility research is to practice

If you are informed about the legibility research that has been done, why it has been done, how it has been done, and what the outcomes are, you are in a position to evaluate its contribution to your design thinking and practice. I therefore encourage you to read on.

Chapter 2

How we read

Rationale

You may question why it is necessary for graphic or typographic designers to know about the mechanics of reading, which would seem to be the responsibility of scientists, particularly psychologists [1] . In order to know what makes a text more legible, we could limit ourselves to finding out about the results of specific legibility studies. But to understand why something may be harder to read, we need to have some knowledge of how we read, in particular the early visual perceptual processes in reading. This stage of identifying letters and words has been described as the perceptual processes of pattern recognition, and this is where design decisions (determining the visual characteristics of letters or lines of text on a page or screen) can have an effect. The written word has been described as a visual object and a linguistic entity (Grainger, 2016). Designers may not be so concerned about the linguistic entity but considering words as visual objects seems key to the role of a typographer or graphic designer. As a psychologist, I am interested in how specific typographic variables affect how we read. I believe this is also very useful information for designers.

Eye movement

A lot of our knowledge of the reading process comes from studies of eye movements. Our eyes do not move along lines of text in a smooth gradual way. Instead, our eyes make ‘saccades’, which are very quick jumps from one point to another, typically jumping 7 to 9 letters (Figure 2.1). During these movements we have no vision; the vision takes place in the pauses or fixations between saccades. These tend to last about 200 to 250 msecs (a quarter of a second). The time spent in pauses is about 90% of the time. When we get to the end of a line, we make a return sweep to the beginning (or close to the beginning) of the next line. If we do not read something properly, we make a 'regression' which is where we go back to an earlier point. When we make these saccades, we position our eyes so that part of the text falls on the area of maximum acuity on our retina; this area is called the fovea (see Figure 2.2). At normal reading distances about 6 or 7 letters fall onto the fovea; adjacent to this is the parafovea and peripheral vision. We have an area of effective vision during a fixation, sometimes referred to as the 'perceptual span', and we make use of letters surrounding the 6 or 7 letters. When reading from left to right, the span typically covers 3 or 4 letter spaces to the left of fixation and 14 or 15 to the right (see Figure 2.3). However, this is not fixed as, for example, beginning readers have a smaller span and text difficulty reduces the span (Rayner, 1986).

Question: Why might studies of eye movements be a good way of finding out how we read? Are we able to report on our own reading?

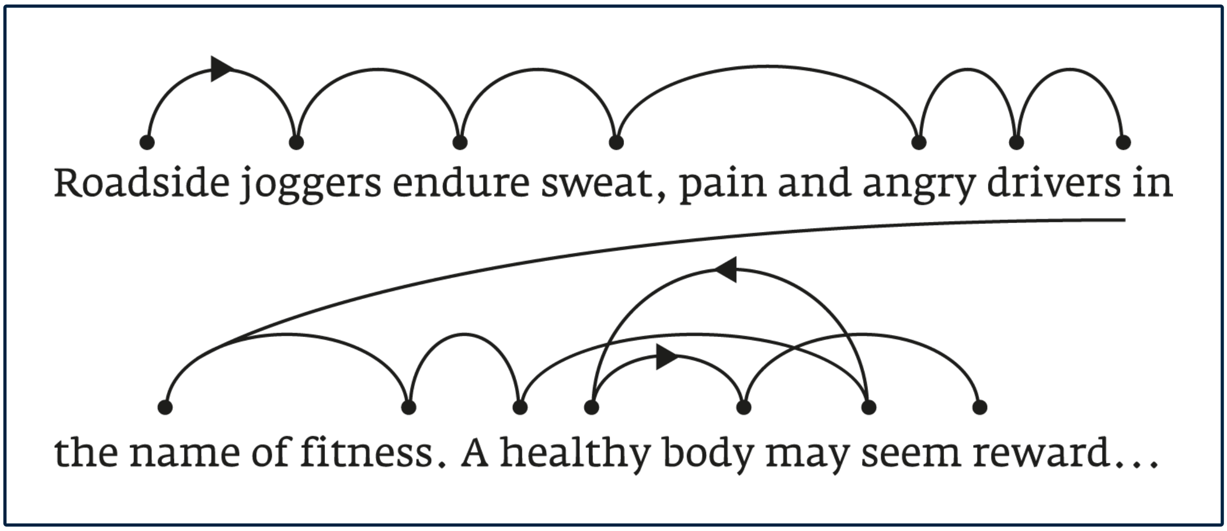


Figure 2.1: A typical pattern of eye movements indicating where on a word our eye fixates (red dots, usually towards the beginning of a word), the length of saccades (jumps), the return sweep from near the end of the first line to near the beginning of the next, and a regression back to the word ‘healthy’ followed by an additional fixation on ‘body’. Diagram based on Larson (2004) and Rayner and Pollatsek (1989, p116).

How do we recognise words?

There is broad agreement amongst reading researchers that word recognition is letter-based. What we are doing in the pauses or fixations is identifying letters and these are combined into words.

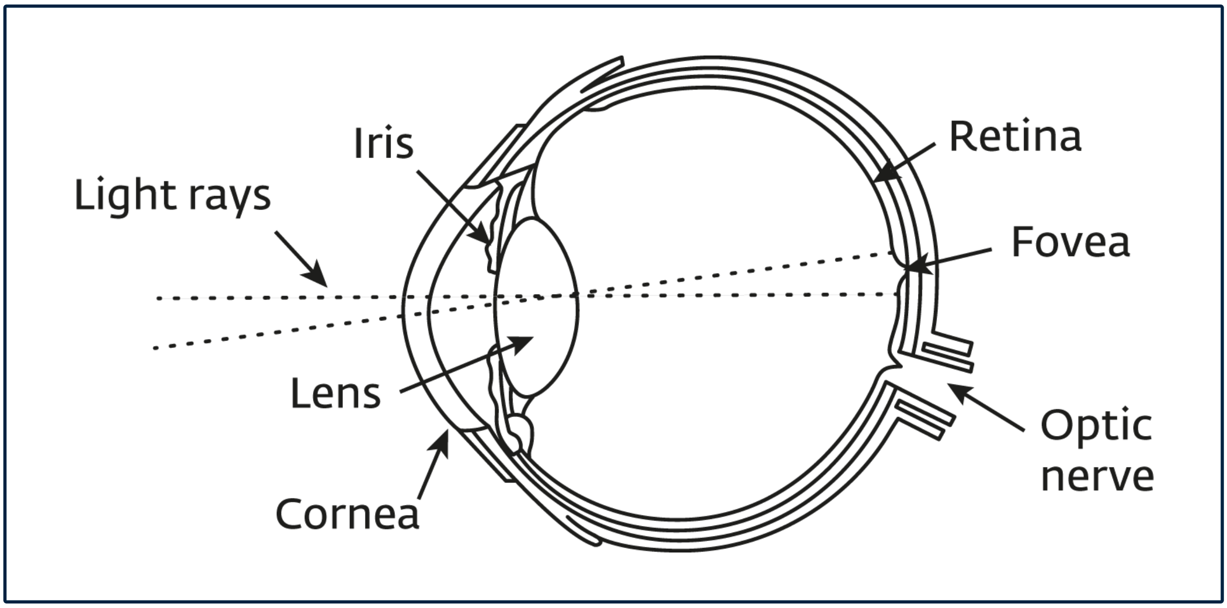


Figure 2.2: Anatomy of the eye showing the retina (at the back of the eye) and the area of the retina with maximum acuity (fovea).

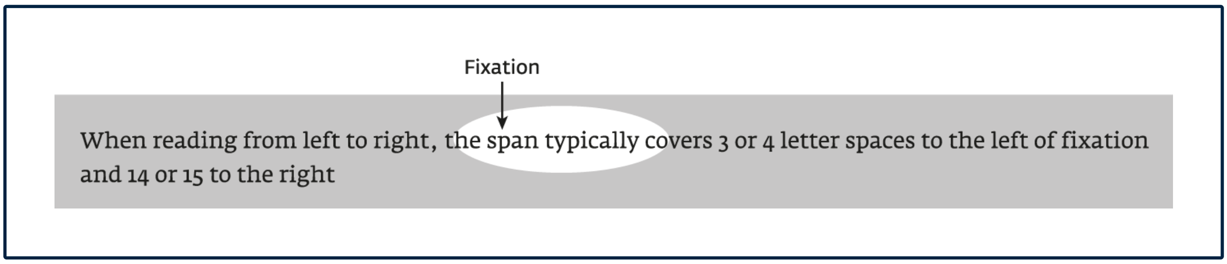


Figure 2.3: An example of the perceptual span and fixation point of skilled readers.

Word shape re-examined

Figure 2.4: Word shape for lower case (small letters) and upper case (capitals) with ascenders and descenders creating an outline in lower case.

However, many texts on typography refer to the use of word shape information, suggesting that we recognise words from their outline shape, e.g. the pattern of ascenders and descenders (see Figure 2.4). This comes from an outdated model, originally proposed in 1886 by a psychologist, James Cattell. Classic texts connected with legibility include references to word shape, as this was probably the current, or reasonably current, thinking based on psychological literature at the time of publication. Spencer wrote: ‘Perception in normal reading is by word wholes…’(Spencer, 1968, p20). Unfortunately, this view is perpetuated in more recently published literature making it important that we critically evaluate what we read.

At an Association Typographique Internationale conference in September 2003, Kevin Larson (a reading psychologist working in Microsoft Corporation’s Advanced Reading Technology Group) spoke of the significant discrepancy between recent psychological models of reading (supported by evidence) and typographers’ beliefs and understanding. Panel 2.1 based on Larson (2004), explains where the support for word shape came from.

?

Question: Why do you think the belief that word shape is important in reading persisted for a long time and is still held by some people?

Panel 2.1: Explanation of where the support for word shape came from

> *Why did the outline formed by the word shape seem to be convincing as an explanation for how we recognise words?*

The first four sources of evidence for word shape are provided by Larson (before he knocks them down). The fifth comes from an Internet text and the source is not entirely certain.

1. Cattell (1886) discovered the ‘word superiority effect’ where word naming is easier than letter naming. He found out by presenting either letters or words to participants for a short time (5–10ms) and more words were accurately recognised than letters. This led to the logical assumption that written words are identified using holistic word shape information.

> *BUT this effect can be attributed to regular letter combinations, rather than word shape.*

2. Further support for word shape appeared to come from the finding that lower case text is read faster than all upper case text. The outline shape of lower case appears to be much more informative (Figure 2.4).

> *BUT one explanation for this is that it is a practice effect because we are used to reading lower case and are therefore more proficient and read it faster (see Chapter 5: Upper versus lower case)*

3. Proof reading errors can be more easily detected if the error changes the word shape:

`correct spelling: many`

`incorrect spelling with same word shape: manj`

`incorrect spelling with different word shape: mano`

> *BUT these results were found to be caused by changes to letter shapes, and not word shapes. In the original studies (Haber and Schindler, 1981; Monk and Hulme, 1983), the word and letter shapes were confounded, meaning that changes to the word shapes also changed letter shapes. It was therefore not possible to separate the two explanations. A subsequent study (Paap, Newsome, and Noel, 1984) changed word shape and letter shape independently and identified that errors that retain the same letter shapes are more difficult to detect in proofreading than errors where the letter shapes are different. The word shape is not relevant.*

4. It is more difficult to read text in AlTeRnAtInGg case than not in alternating case. The explanation for this is that we lose the familiar word shape when alternating case.

   > *BUT this effect is also true for pseudowords that we have not encountered before and therefore the word shape would not be familiar.*

5. Another demonstration that seems to support word shape was contained in a text circulating on the Internet in 2003 (with my correction).

   > *Aoccdrnig to a rscheearer*[2] *at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.*

   Which reads as:

   > *According to a researcher at Cambridge University, it doesn't matter in what order the letters in a word are, the only important thing is that the first and last letter be at the right place. The rest can be a total mess and you can still read it without problem. This is because the human mind does not read every letter by itself but the word as a whole.*

   Although this claimed to come from Cambridge University, a researcher, Matt David, at the MRC Cognition and Brain Science Unit, Cambridge University, UK tracked down the original demonstration of the effect of letter randomisation to a PhD thesis [3]. As David explains, although some of the content is partially correct, there are also misleading statements.

   Demonstrations, such as this jumbled text, can appear quite convincing because they can be manipulated to support whatever claim is being made – here the claim that this is a readable text. Although it might appear to support the argument for reading words as wholes (and not by letter), another explanation, supported by more recent research and based on letters and not words, is given by Grainger and Whitney (2004). If the jumbled words contain the same letters, small changes can be made to the letter order (called transpositions) and we can still identify the words because there is sufficient information on the correct relative position of letters.

> *Aoccdrnig*
>
> *According*

In fact, we are slowed down by reading sentences with transposed letters – a demonstration cannot reveal such subtle effects. Some transpositions are more problematic than others (Rayner, White, Johnson and Liversedge, 2006). If internal letters are transposed (e.g. reading becomes readnig or redaing) we are not slowed down as much as when beginning or end letters are transposed. Changing the order of letters beginning the words is the most disruptive (e.g. word becoming owrd).

Exercise: Take the jumbled paragraph:

> *Aoccdrnig to a rscheearer at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.*

Can you re-arrange each word so that it becomes problematic to read trying not to move a lot more letters than in the original jumble? Is it still possible to read some words if the beginning and end letters are transposed?

Parallel letter recognition

As skilled readers, we identify individual letters in parallel (simultaneously) rather than sequentially (one after the other). We therefore need to not only work out what the letters are but also their order within words, using word spaces to identify the word boundaries. This information is used to match against stored words to derive meaning and/or sound (pronunciation). See Panel 2.2 for a distinction between silent reading and reading aloud.

Panel 2.2: Theory on distinction between silent reading and reading aloud

Recent work (Grainger, Dufau and Ziegler, 2016) proposes a theoretical framework which incorporates two routes from letters to words:

1. We might go directly from the letters to meaning, a faster but not necessarily entirely accurate route as the precise letter positions may not be known. Instead subsets of letters may be used that help to identify a unique word, and these may not be adjacent letters. If reading silently, this route might be sufficient for our needs.
2. The second route generates sound from the text, using more precise positioning of letters, and we need this route for reading aloud. We read aloud more slowly than silent reading.

In learning to read we start letter-by-letter, one after the other, unlike the parallel processing of skilled readers. The beginning reader identifies individual letters and learns the corresponding sounds of individual letters and combinations. At this stage they will know the letters of the alphabet and have a spoken vocabulary. Their task is to bring these together. This mapping is thought to set up the connections that exist for skilled readers (Grainger, Lété, Bertand, Dufau, and Ziegler, 2012).

The space between letters is also important as letters are less visible when surrounded by other letters. This is referred to as ‘crowding’, and is not specific to letters. The effect of crowding is greater in our peripheral vision, which means we are less able to recognise words further from the fovea. This is due to reduced visual acuity and crowding. Words are recognised from their parts (i.e. letters) and crowding reduces our ability to identify the individual letters as the adjacent letters jumble their appearance.

This also happens with faces. If we look at a face using our peripheral vision, it may be difficult to work out if the person is frowning or smiling. The context of the face hinders our perception. If the mouth alone were shown to us in peripheral vision, without the face context, it would be easier to work out if it was a frown or smile. If we look at someone using our central vision (in front of us), having the whole face is an advantage. Box 2.1 describes the research and hopefully demonstrates this effect.

Word context

A large amount of research into how we read has used isolated letters and words that are not in the context of sentences. These studies might therefore be criticised for using artificial test material which does not reflect 'normal reading'. (Chapter 4 will discuss the reasons for the choice of methods in more detail). We usually read words in sentences and this context can help us to predict what word may follow. The description of crowding above has also shown that context can have a negative effect (in peripheral vision). A psychological study has looked at the relative contribution of letters, words, and sentence context to how fast we read (Pelli and Tillman, 2007). They found that letters contribute most to reading rate (62%); words contribute only 16% and sentence context contributes the remaining 22%. Word shape therefore plays a very small part in reading [4] . The research is described in Box 2.2.

?

Question: How easily can you read the following? Why is this more difficult than the demonstration in Panel 2.1 from the text circulating on the Internet? (Reading Box 2.2 may help)

That and frist word Uinervtisy at the Ltteers thing rscheearer pclae to are a the is mttaer Cmabrigde aoccdrnig it in lsat the deosn't oredr olny what ltteer rghit iprmoetnt at what be a.

Box 2.1: Details of 'face inferiority' effect

A rather interesting series of studies asked the question as to whether faces are processed like words. Do we identify a word or face as a whole or by its parts? By now you will know the answer to how we recognise words: by identifying the individual letters.

The researchers (Martelli, Majaj and Pelli, 2005) include a demonstration of the effects which I hope you will be able to experience for yourself in Figure 2.5. They refer to these as the 'word inferiority' and 'face inferiority' effects. You may remember that Cattell invented the 'word superiority' effect (Panel 2.1). Here we have more evidence that argues against word superiority and reading by word shape.

In Figure 2.5, at the top you have the word inferiority effect. If you focus on the square in the middle and try to identify the middle letter on the left, you may find that it is difficult to do. If you again focus on the square and try to identify the letter on the right, it should be much easier to do. Similarly for the face, focus on the square in the centre and see if you can tell whether the face is smiling or frowning. Then do the same for the mouth on the left. You may find that it is much easier to tell whether you see frowning or smiling when there is no context (the face).

<img src="/assets/illustrations/FIG-2-5-TODO.jpg" alt="TODO - missing image"

Figure 2.5: These images demonstrate the effect of context or crowding on letters and faces. You need to use your peripheral vision by focusing on the square in the middle. On the left, the letter or mouth is shown in the context of other letters or the face. This context which makes it more difficult to accurately perceive them compared to the items out of context shown on the right.

Box 2.2: Details of

contributions to reading from letters, words, and

sentences

The way the researchers calculated how much letters, words and sentences contribute to reading was to systematically remove each source of information. Sentence information was removed by changing word order (Figure 2.6a); word shape information was distorted by alternating case (Figure 2.6b); letters were substituted for similar shapes so that the whole word shape was preserved (Figure 2.6c). The reading rates at which participants achieved an accuracy of 80% (i.e. 20% of words were incorrectly reported) were measured for all combinations of these three methods. They didn't just remove one at a time, but also removed two cues (e.g. just having word information, removing letters and sentence information). An example from their paper:\ \ a ard ct nocm ct fbet ba

If we add back the letter and sentence information, it reads as:

at the end of the room a

?

Question: Which of Figures 2.6a, 2.6b, 2.6c looks hardest to read and understand? Which looks easiest?

a) Losing sentence information

different random in sentences order Words from

b) Losing word information

ThEsE wOrDs InTeRwEaVe LoWeR aNd UpPeR cAsE

c) Losing letter information

Tbis sartcrec suhfitufes similen lctfans

Figure 2.6: How sentence, word and letter information was removed by Pelli and Tillman (2007).

> *Did you think that losing letter information (c) made reading hardest and losing word information (b) was easiest to read? If so, your answers correspond to the results of Pelli and Tillman.*

Identification of letter features

Figure 2.7: Font used to create words in Rumelhart and Siple (1974) and still used in models of reading.

Given the importance of identifying letters, quite a lot of research has looked into what features of letters we use to distinguish one letter from another. However, models of reading have assumed that the particular font will not affect the basic results (McClelland and Rumelhart, 1981, p383). Many models use a font with straight-line segments, created by Rumelhart and Siple (1974) which disregards typical letter shapes (see Figure 2.7). However, there is now a greater focus on letter perception by psychologists which must be good for typographers. The outcomes of these studies are described in Chapter 5 where they are combined with research from a design perspective.

Reading different typefaces and handwriting

The research on letter features looks for characteristics that are shared by all letter `a`s and letter `b`s etc., such as mid segments or stroke terminals (see Chapter 5: Letter features). A skilled reader can recognise most letters quickly regardless of the visual form, which can mean the font, case (capital letters and small letters), or style of handwriting,

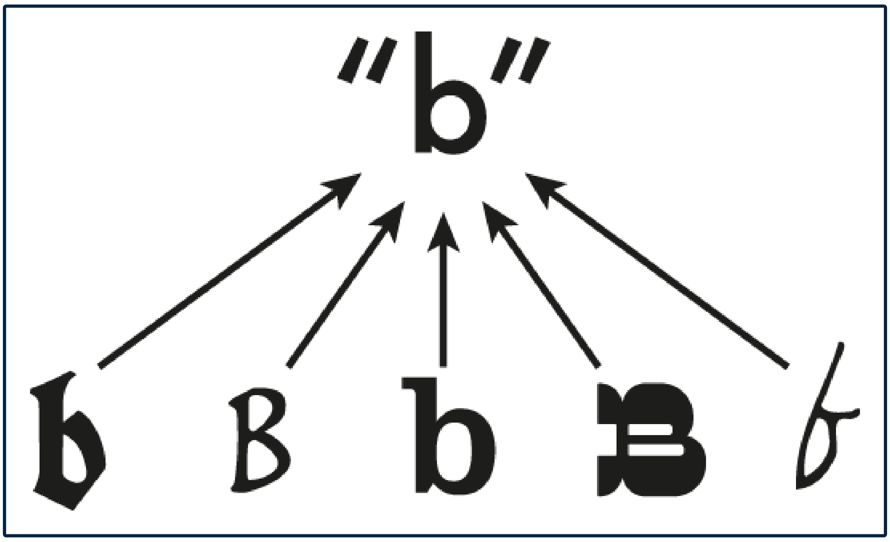


Figure 2.8: Readers identify the letter b even though it can take many shapes and sizes.

Despite these differences in the visual forms of the same letter, we can easily identify letters, recognising them as representing the same character. We are creating abstract letter identities (Grainger, Rey and Dufau, 2008), where the letter is identified as `a` or `b` irrespective of font, size or case (Besner, Coltheart and Davelaar, 1984). Figure 2.8 illustrates this mapping of different forms onto a single representation. How we do this, and identify letters despite their different forms, was proposed by a psychologist about 30 years ago (Sanocki, 1987, 1988). He referred to this as 'font tuning'.

It is often assumed that once we have converted to an abstract letter identity, we no longer retain knowledge of the visual form, because this is not essential to reading. Some exceptions to this are when we wish to:

- identify the typeface (something that typographic and graphic designers may wish to do)
- recognise whose handwriting we are looking at
- identify brand names and corporate identities

Panel 2.3 provides a little more detail of font tuning and research which looks at how we recognise letters using neuroscience techniques.

Panel 2.3: Font tuning and neuroscience research

Essentially, we can use the characteristics of a particular font or style of handwriting to help us identify letters and convert from the visual form containing specific properties of that font (variants of the same character) to abstract representations (invariant forms). This happens extremely quickly from recognising the form as a letter (100–200msecs) to recognising the specific letter (120–180msecs) to abstracting the invariant form (220msecs) to indicating recognition (after 300msecs). (Rey, Dufau, Massol and Grainger, 2009 cited in Thiessen, Kohler, Churches, Coussens, and Keage, 2015).

Some recent research using neuroscience tools and methodologies suggests that the visual form of letters may have an effect after abstract information is extracted (Keage, Coussens, Kohler, Thiessen and Churches, 2014), even though we don’t necessarily need to remember a typeface.

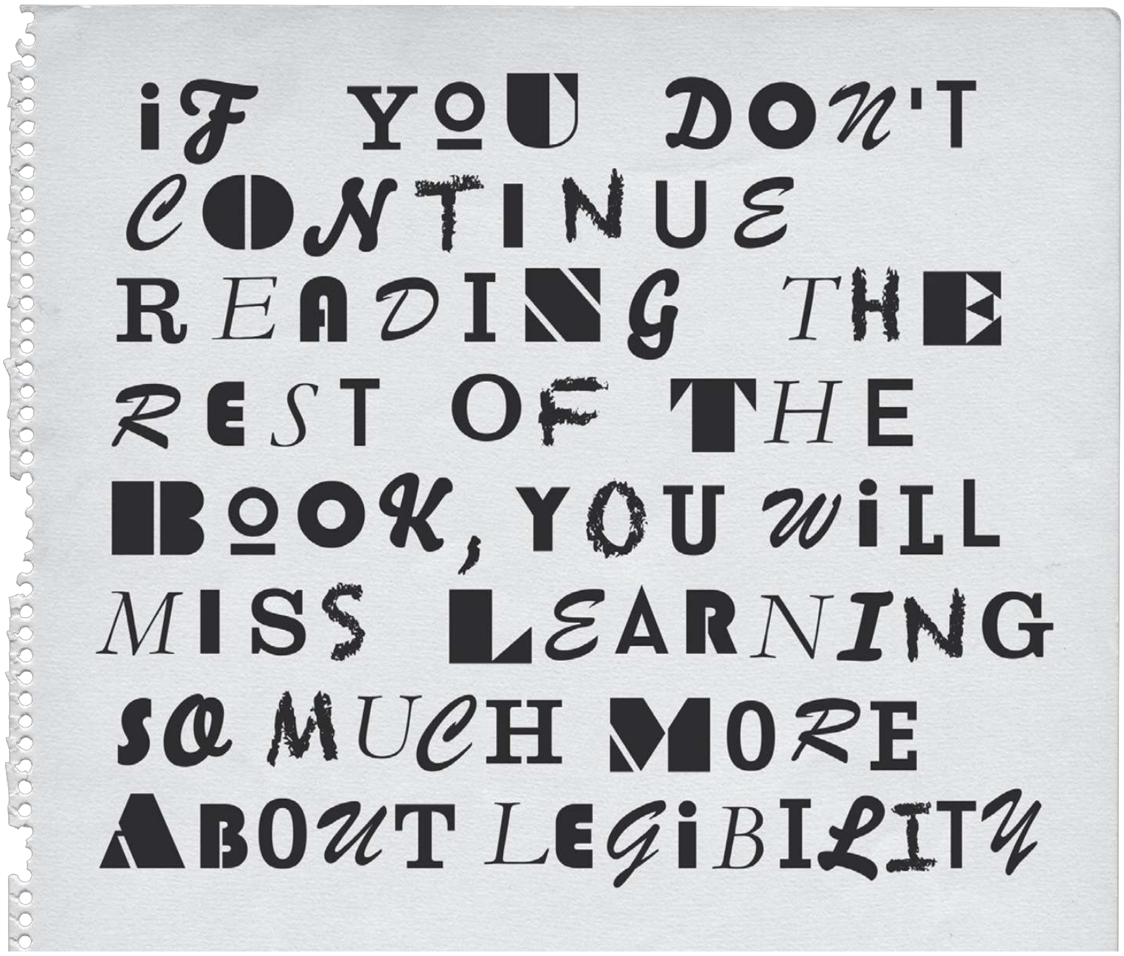


Figure 2.9: A digital 'ransom note' simulating the practice of cutting out letters from different newspapers as a way of avoiding your handwriting being recognised and therefore ensuring anonymity. There is no point in doing the same in a digital environment yet ransom note fonts exist. This creates a challenge for font tuning and letter identification as we need to re-tune letter by letter.

?

Questions: why is handwriting usually harder to read than print, based on what you have learned about how we read? Think about (i) individual characters; (ii) relationship between different characters.
Here's a clue: Why might a ransom note be more difficult to read than normal text? (Figure 2.9)

Summary

Typographers and graphic designers were led to believe that we read by identifying words from their outline shape. This was once the view held by psychologists, but research improves our understanding and it is important to update our knowledge. We know a lot about reading from:

- monitoring eye movements
- using sophisticated techniques to see which parts of letters we use to differentiate letters
- working out how sentence and word information contribute in positive (providing context) or negative (crowding) ways

There is a greater interest developing among scientists in looking at different visual forms, not just assuming all letters are equal so the font or case doesn’t matter. Recent psychological research is demonstrating a greater sensitivity to typography which will be of great benefit to designers. This is described further in the next chapter.

Chapter 3

Perspectives on legibility

Historical perspective

What are the origins of legibility research? A few landmarks are worth reporting as they formed the foundations for subsequent research and are frequently cited. Various writers (Spencer, 1968; Rehe, 1979; Tinker, 1965) propose that scientific legibility research began with Javal around 1880, a French ophthalmologist who studied eye movements in reading. In particular, Huey wrote a book on *The psychology and pedagogy of reading* [1] in 1908 (reprinted 1968) which credits Javal with discovering the pattern of eye movements in reading (described in Chapter 2).

Closer examination of the original sources by psychologists reveals a different story (Wade, Tatler and Heller, 2003). In 1879, a physiologist Hering first discovered that reading involves saccadic eye movements and Lamare in 1892, a colleague of Javal, noted the jerky, rather than continuous, movements. Dodge (a psychologist) was able to develop a photographic technique which enabled more accurate measures of the speed of saccades and the length of fixations. These developments were the start of eye movement recording technologies.

Shortly after the discoveries concerning eye movements, Cattell (1886) claimed to have found the word superiority effect (as mentioned in Chapter 2). Other work around this time that is often cited includes:

- which letters are more legible than other letters (Sanford, 1888; Roethlein, 1912)

- a review of early legibility research by Pyke (1926)

Research directions

Although the above research looking at visual mechanisms in reading began in the late nineteenth century, visual science moved away from applied research and therefore lost a connection with legibility research. The psychology of reading became the province of cognitive psychology, education, and psycholinguistics with less interest in typographic and graphic aspects of text. From around 1980, computational models of reading were developed aiming to simulate the recognition of words through mathematical modelling with computers.

An extensive programme of legibility research was conducted by Tinker and his colleagues, which did not attempt to explain the underlying visual mechanisms for the results. To some extent, this reflected the state of knowledge at that time, and separation of theoretical and applied research. 'Theoretical', 'basic' or 'pure' research has the aim of investigating the visual processes involved in reading whereas applied research aims to evaluate which typographic solutions are better for reading. Tinker and colleagues carried out numerous experiments between the 1920s and 1950s which provide a substantial body of findings. Given our current knowledge of how we read, and more recent interest in visual processing relevant to legibility (Legge, 2007, p108), these results from traditional legibility research can now be more thoroughly evaluated and interpreted.

Reviews of legibility research summarise what is known at the time by discussing research published by others, and which might also include research by the author (e.g. Ovink, 1938; Tinker, 1963, 1965; Zachrisson, 1965; Spencer, 1968; Foster, 1980; Reynolds, 1984; Lund, 1999; Dyson, 2005; Beier, 2012). These can be useful texts for gaining an overview of research findings, which should again be critically evaluated.

Design perspective

Legibility research has not typically been carried out by designers, as they are unlikely to have a detailed knowledge of scientific method, normally acquired over the course of a psychology degree. However, designers have views on what should be researched and how it should be researched. These views do not necessarily fit with scientific approaches to research. The objectives of the research usually differ across disciplines and these can determine the methods used.

The primary method used by the psychologist concerned with reading research is empirical experimentation (Rayner and Pollatsek, 1989, p8). Within the psychology of reading there are various perspectives including cognitive and linguistic constructs (e.g. Kintsch and van Dijk, 1978), perceptual factors related to text legibility (e.g. Tinker, 1963, 1965), and the nature of visual information processing in reading (e.g. Legge, 2007).

A designer's primary motivation is to make text more legible, i.e. easier to read, and also aesthetically pleasing — a practical approach. Designers often express dissatisfaction with certain aspects of empirical research, typically carried out by psychologists without consulting designers. Many of the reasons, raised in the past and still debated, are reviewed and summarised by Lund (1999) and Beier (2016). Below I list a few of these criticisms (from a design standpoint) of legibility research and then try to address these criticisms. These focus on the value and the relevance of the research to design practice. Issues which relate to the research methods are dealt with in the next chapter.

- A fundamental position opposing legibility research argues that research inhibits creativity, which seems to assume that the outcomes of research must be implemented. Poynor (1999, p14) suggests that designers view the scientific approach as an opponent of the creative process, and do not wish to be led by psychologists.

- In 1970, Merald E. Wrolstad, editor of the *Journal of Typographic Research* speaking about the misunderstandings between disciplines, pointed out that research will never replace the creative designer; not everyone wants to be a typographer. Research should be regarded as a source of ideas (Wrolstad, 1970, p37–38). A solution may therefore be for designers to familiarise themselves with legibility research and then consider whether it is helpful to them in informing design practice. This does require researchers to present their work in an accessible form, and ideally including a hint as to how the findings might translate into practice, if relevant.
- Results of much legibility research merely confirm existing practices, based on craft knowledge. Or, results are contradictory and therefore of little use.
- If empirical research supports existing practice, this can be seen as positive and reassuring. Where there are discrepancies, either between existing practice and research or among different studies, there is more cause for concern. Such issues may be resolved with further well-planned studies.
- Researchers naturally focus on areas and questions which interest them (Wrolstad, 1969, p116), which may not be particularly relevant to design practice.
- Some research (such as the studies in Chapter 2) is not intended to have direct relevance to design practice. Pure research falls into this category. When reading about a study, it is important to work out why the research was done, which should be found in the introduction. The affiliations of the authors may also indicate their discipline if attached to a university or research centre and these are commonly included in the journal article (see Figure 3.1).

Does print size matter for reading? A review of findings from vision science and typography

Gordon E. Legge Department of Psychology, University of Minnesota, Minneapolis, MN, USA

Charles A. Bigelow School of Print Media, Rochester Institute of Technology, Rochester, NY, USA

Figure 3.1: Part of the first page of an article by Legge and Bigelow (2011) published in the *Journal of Vision* listing the affiliations of the two authors and therefore indicating their contrasting disciplines. In this particular instance, the title also makes this very clear.

- Researchers should not look for universal truths but aim to provide practical decision-making guidance (MacDonald-Ross and Waller, 1975, p77).
- Practical guidance may require testing a specific version of a design, which is better described as formative evaluation or user/usability testing, rather than legibility research. The results will apply to the particular situation and will not be generalizable. (See Chapter 4: Different types of testing and research).
- Researchers chose topics that are easy to investigate in the laboratory (MacDonald-Ross and Waller, 1975, p76)
- This may have been true in 1975 and earlier, but I do not believe that this is currently the case as very sophisticated techniques have been developed. However, most of the research is done under controlled conditions (i.e. a 'laboratory setting'). This is covered in more detail in Chapter 4.

?

Questions: If you were asked to debate the value of legibility research, which side would you prefer to argue: for or against? Which of the points above do you think are the strongest? Can you add any other points?

Combining resources across disciplines

Collaboration between people with diverse backgrounds and expertise can lead to mutual understanding of the important and different contribution that comes from another discipline. Engaging in discussion can help us understand the other's viewpoint which should make us less dismissive of alternative perspectives.

Exercise:

Fernand Baudin (1918–2005), a Belgian book designer, author, typographer, and teacher, objected to Tinker's description of typographers as aesthetes when reviewing the book 'Bases for effective reading' (Baudin, 1967). I have extracted excerpts from the pages listed by Baudin, which I think are the parts in the book that he references.

> *Consider whether you think Baudin was justified in being upset by the statements (quotations) below from Tinker's book.*
>
> *Do you think Baudin was right in interpreting the statements as: '...all typographers en bloc, whether expert or not, are presented merely as introspective aesthetes deserving, on the whole, of contempt' (p204–205).*
>
> *Is Tinker criticising typographers with these statements?*
> *Is it an insult to be concerned with aesthetics?*

- 'Before scientific research, printers and type designers were concerned mainly with the esthetic appearance of the printed page.' (p115)
- '...the dominant guides to typography until rather recently were esthetics, economy of printing, and traditional practice.' (p125)
  'The subjective opinions of type designers and typographers as to legibility of letters prevailed throughout the nineteenth century and have carried much weight even up to the present day.' (p125)
- 'This practice continues even though many typography "experts" consider that italic

type is far less legible than regular Roman lower case'. (p135)

- 'Although some designers may have a strong esthetic objection to boldface for headings, this does not mean that readers react the same way.' (p136)
- 'The strong belief that generous margins will increase legibility agrees with the opinions of most "experts" expressed between 1883 and 1911 (Pyke, 1926) (p183)
- 'While there is an "average" consensus, printing practice in use of margins in individual books varies greatly (Paterson and Tinker, 1940). Whether this is motivated by an attempt to produce a more pleasing page or by an unconscious departure from the 50 per cent rule, or both, is uncertain.' (p183) [The 50 per cent rule refers to the general practice of publishers to use 50 per cent of a page for margins (Tinker, 1965, p182)].

If we look at the above quotations from a more neutral perspective, we might suggest that Tinker was wishing to make a clear distinction between scientific research (admittedly, his own) and the craft knowledge of typographic experts. The comments are not limited to aesthetics as legibility is included. However, it is unfortunate that Tinker uses quotation marks around the word "experts" which might be seen as an ironic comment.

Fortunately, we have moved on from Tinker and recognise that combining skills and knowledge across disciplines can result in more relevant and robust research. An example of an excellent collaboration between vision scientist and type designer is the article illustrated in Figure 3.1 and described in Panel 3.1. Other examples of collaborations where design expertise combines with scientific knowledge include:

- *James Hartley and Peter Burnhill:*
  Burnhill was a teacher of typography (1923—2007) who engaged in a long collaboration with James Hartley, a psychologist at Keele University, UK who is still a very active researcher and writer. This duo explored how typography can support readers' use of texts through clearly displaying the structure of the texts (e.g. use of space).
- *Robert A. Morris, Kathy Aquilante, Charles Bigelow, and Dean Yager:*
  In 2002, vision scientists (Aquilante and Yager) combined with a mathematician working in computer science (Morris) and type designer (Bigelow) to look at how

serifs affect reading on screen.

- *Owen Churches, Scott Coussens, Hannah Keage, Mark Kohler, and Myra Thiessen:* Thiessen is a designer with knowledge of how to conduct experiments; all the other members of the team are neuropsychologists and together they have looked at how the brain processes typography using EEG (electroencephalography) technology. Their research is mentioned in Panel 2.3.

Panel 3.1: How type is measured by typographers and vision scientists

Gordon Legge and Charles (Chuck) Bigelow explain the different way that type is measured in the two disciplines: typographers describe the physical size of type on the page or screen, commonly in points; vision scientists combine the physical size and the viewing distance of the reader, referred to as the angular size or visual angle (see Figure 3.2). The reason for this (apparently) more complicated measure is that the image on our retina will be smaller if we are further away from the type and the retinal image is what is most relevant (see Chapter 2 for details of the eye).

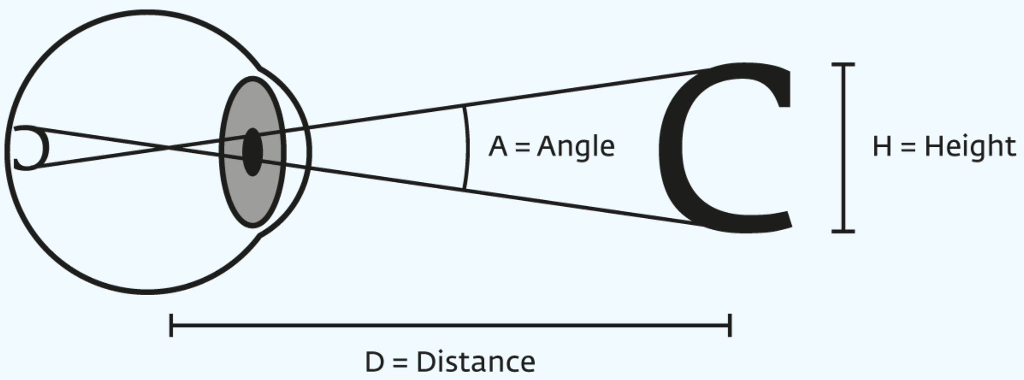


Figure 3.2: Diagram showing how the visual angle of an object (in this case a letter) is measured. The formula for calculating the angle is Visual angle = 2arctan(height of object/2 x distance).

Summary

Legibility research started with eye movement research over 100 years ago. Some of these discoveries and writings are still valid today, whilst others have been superseded as research has enabled more precise measurements and a larger body of knowledge has developed.

Legibility straddles disciplines, broadly science and design, and in the past, this has caused tension due to different objectives and, at times, insensitive appraisal of other perspectives. As more collaborations are developed, richer, more relevant, and more robust research findings emerge to inform typographic practice.

Chapter 4

What is measured and how

Different types of testing and research

A distinction can be made between testing that is carried out as part of the design process and testing on finished products.

- *Formative evaluations*, i.e. before finalising the design, can inform design decisions by either detecting problems with some aspects of a single design (e.g. type is too small) or indicating which of two or more versions is easier to read.
  This form of testing is described as *diagnostic testing* when pinpointing specific problems, and is ideally used as part of an iterative design process. Having detected a problem, this is resolved and then re-tested.
- *User testing* or *user research* compares different versions and this may be carried out as a formative evaluation, to determine which version to develop further.
- If user testing is carried out as a *summative evaluation*, i.e. testing the final product, the results may provide recommendations for the design of future similar products. However, this practical guidance will be limited if it is not possible to determine why one version was better than another.
- *Research studies* make comparisons between different versions whilst controlling how they vary. From these results, it should be possible to say, for example, which typographic variable affects speed of reading. The research is therefore generalizable to other design situations and can be considered robust research, if carried out appropriately.

Question: Consider whether you have used a formative evaluation as part of your design process. For example, have you asked colleagues or friends for feedback about aspects of your design?

Challenges

Key criteria

The methods used for the first three types of testing above can be less formal than those used for research studies. In some circumstances, it may be unnecessary to meet all of the criteria listed below, or they may be less relevant. Nevertheless, it is helpful to know what are the main challenges to carrying out robust research that will be of value and relevant to both researchers and designers.

Although the three criteria are listed separately, they do interrelate. Finding a solution to one challenge may conflict with another so a judgement must be made as to what to prioritise[1].

The key criteria in designing a study are:

- Sensitivity: finding a method to measure performance of some aspect of reading that is sensitive enough to pick up differences when typography is varied.
- Reliability: ensuring that the results you get are repeatable. If you were to do the same study again, would you get the same outcome? One solution is to increase the amount of data collected. You can do this by using a sufficiently large number of participants in the study and, where practical, giving participants multiple examples of each condition of the experiment. These requirements present their own challenges which are to find enough participants and to fit the experiment into a reasonable length of time.
- Validity: determining that the study measures what it is intended to measure. Of most relevance to legibility research, and the designer's perspective, is ecological validity, a form of 'external validity'. This describes the extent to which a study approximates typical conditions and is also referred to as 'face validity'. In our context, this can mean a natural reading situation and suitable reading material. Another form of validity is 'internal validity' which describes the relationship between the outcomes of the study and the object of study. This is explained further below.

Reading conditions

Ecological validity is not only a concern of design practitioners but also of psychologists doing applied research. However, reading situations in experiments are frequently artificial and do not resemble everyday reading practice. As mentioned in Chapter 2, research has often looked at individual letters or words, rather than reading of continuous text. The letter or word is often displayed for only a short time and the participants in the studies may be required to respond quickly. Context is also removed which means:

- If testing individual letters, there are no cues from other letters which might help identification. Panel 4.1 provides an example of how the stylistic characteristics of a particular font, or style of handwriting, may help us identify letters.
- If testing words, there is no sentence context.

Panel 4.1: Stylistic consistency within a font

In Chapter 2 I described how we are able to read different visual forms of the same letter. However, in reading normal texts (rather than ransom notes) we do not need to switch between lots of fonts. We read paragraphs of text where the stylistic characteristics of the letters provide cues to what other letters will look like (see Panel 2.3: Font tuning). Repeated parts of letterforms, such as curves, are stylistically consistent in shape, weight, etc. (see Figure 4.1)

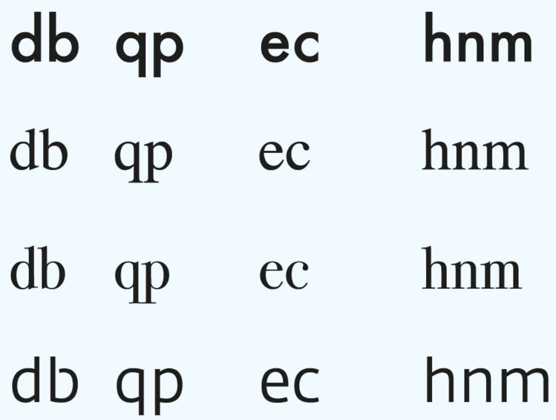


Figure 4.1: Groups of letters in 4 different fonts showing the stylistic consistency within the font. Type designers group letters with similar forms to apply consistent stylistic attributes.

Clearly these are not everyday reading conditions, but there are compelling reasons for carrying out a study in this way. These techniques can be necessary to detect quite small differences in how we read because skilled readers can recognise words very quickly (within a fraction of a second). Any differences in legibility need to be teased out by focusing on a part of the reading process and making that process sufficiently difficult to detect change. This is a way of making the measure sensitive (one of the three criteria described above), but at the expense of ecological validity. Although some research does use full sentences and paragraphs, these may not always reveal differences or may be testing different aspects of the reading process.

Designers, in particular, can also be critical of studies measuring speed of reading claiming that how fast we read is not an important issue for them. Speed of reading, or speed of responding to a single letter or word, are also techniques used to detect small differences, and may be used because they are reasonably sensitive measures. It is not the speed itself which is important but what this reveals, e.g. ease of reading or recognition.

Material used in studies

Another criticism relating to artificial conditions in experiments is the poor choice of typographic material, e.g. the typeface or way in which the text is set (spacing, length of line etc.). The objection to such material is that designers would never create material in this form and therefore it is pointless to test; the results will not inform design practice. In some cases, there is no reason for the poor typography of material used in a study, except the researcher's lack of design knowledge. The researcher may not be aware that it is not typical practice. In other cases, the researcher may need to control the design of the typographic material to ensure that the results are internally valid. If I am interested in the effect of line length I could:

- Compare two line lengths and also vary the line spacing (see Figure 4.2). Experienced typographic designers increase the space between lines when lines are longer. But if I set the text in this way I cannot be sure if the line length or the spacing, or both, have influenced my results. The line spacing is a confounding variable.
- Compare two line lengths and not vary the line spacing (see Figure 4.3). But designers will say that they would not create material which looks like this.

Another criticism relating to artificial conditions in experiments is the poor choice of typographic material, e.g. the typeface or way in which the text is set (spacing, length of line, etc.) The objection to such material is that designers would never create material in this form and therefore it is pointless to test; the results will not inform design practice. In some cases, there is no reason for the poor typography of material used in a study, except the researcher's lack of design knowledge. The researcher may not be aware that it is not normal practice. In other cases, the researcher may need to control the design of the typographic material to ensure that the results are internally valid.

Another criticism relating to artificial conditions in experiments is the poor choice of typographic material, e.g. the typeface or way in which the text is set (spacing, length of line, etc.) The objection to such material is that designers would never create material in this form and therefore it is pointless to test; the results will not inform design practice. In some cases, there is no reason for the poor typography of material used in a study, except the researcher's lack of design knowledge. The researcher may not be aware that it is not normal practice. In other cases, the researcher may need to control the design of the typographic material to ensure that the results are internally valid.

Figure 4.2: Comparison of line lengths of around 50 and 100 characters per line (cpl) with adjustments to line spacing. The shorter line length is 10 point type with 12 point line spacing; the longer line is 10 point type with 14 point line spacing.

Another criticism relating to artificial conditions in experiments is the poor choice of typographic material, e.g. the typeface or way in which the text is set (spacing, length of line, etc.) The objection to such material is that designers would never create material in this form and therefore it is pointless to test; the results will not inform design practice. In some cases, there is no reason for the poor typography of material used in a study, except the researcher's lack of design knowledge. The researcher may not be aware that it is not normal practice. In other cases, the researcher may need to control the design of the typographic material to ensure that the results are internally valid.

Another criticism relating to artificial conditions in experiments is the poor choice of typographic material, e.g. the typeface or way in which the text is set (spacing, length of line, etc.) The objection to such material is that designers would never create material in this form and therefore it is pointless to test; the results will not inform design practice. In some cases, there is no reason for the poor typography of material used in a study, except the researcher's lack of design knowledge. The researcher may not be aware that it is not normal practice. In other cases, the researcher may need to control the design of the typographic material to ensure that the results are internally valid.

Figure 4.3: Comparison of line lengths of around 50 and 100 characters per line with no adjustments to line spacing. Both line lengths use 10 point type with 12 point line spacing.

In these two examples, there is a conflict between the internal validity, ensuring that the study is planned correctly, and ecological validity. See Panel 4.2 for further detail of experiment design.

?

Question: Are you convinced by the reasons I have given for using unnatural conditions and test material? If not, what are your concerns?

Panel 4.2: Explanation of interacting typographic variables in

psychology experiments

Typographic and graphic designers learn to make decisions about type size, line length, and line spacing in relation to each other. These typographic variables are considered to be inter-related. In psychology experiments, this inter-relationship can be demonstrated by finding interactions between the variables. In the example of line lengths and line space (Figures 4.2 and 4.3), if the type size remains constant, we might expect to find that optimal legibility for a longer line length has larger line space and optimal legibility for a shorter line length has a smaller line space.

In Figure 4.4 I have plotted some data from Paterson and Tinker, reproduced in Tinker (1963, p95). The study used 10 point type and I have selected three line lengths (around 40, 54 and 90 characters per line) with line spacing starting from 10 point and increasing to 11, 12 and 14 point. At all three line lengths, 10 point line spacing slows down reading and the line length has very little effect. However, the results regarding optimum combinations of line length and line spacing are not as I predicted above: the optimum line spacing for the longer line length (90 cpl) is 12 point; this is also the optimum for the two shorter line lengths (40 and 54 cpl).

Nevertheless, this is an example of an interaction between line length and line spacing. The effect on reading speed of the amount of spacing depends on the line length. We can see this from the graph as the three lines representing the line lengths are different shapes, indicating a different pattern of data. The consequence of this difference is that if I had chosen not to adjust line spacing as line length varied (as in Figure 4.3), but instead tested all line lengths with 11pt line spacing, I would have concluded that:

- a line length of 40 cpl is read fastest
- 90 cpl is quite a bit slower
- but 90 cpl is read faster than 54 cpl

If I had chosen 12pt line spacing, I would have reached a different conclusion:

- lines of 40 and 54 cpl are read at the same (fast) speed
- lines of 90 cpl are read slower

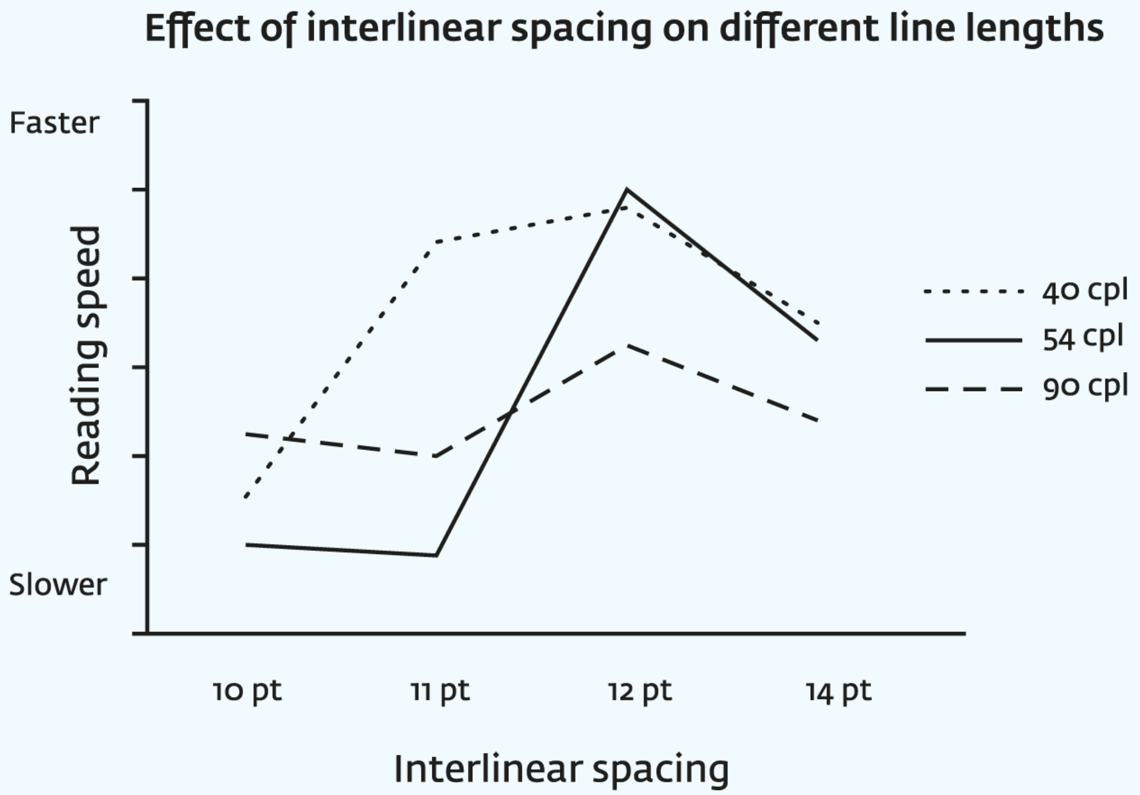


Figure 4.4: Graph showing the relationship between two typographic variables (line spacing and line length) and how this affects legibility measured by reading speed. The graph is based on a subset of data reported in Tinker (1963).

This selective use of data is employed only to illustrate how to translate designers' respect for the relationship between typographic variables into experiment design. It is unwise to regard these specific results as a guide to design practice. Chapter 5 reviews a wider range of research which is more representative of the findings and therefore a better guide.

The data in Figure 4.4 was taken from a huge series of studies in which the experimenters included all combinations of line lengths, line spacing and different type sizes. This scale of testing would not be carried out today as it would not be considered a feasible or efficient approach. Instead, the options would be limited to those shown in Figures 4.2 and 4.3:

- adjusting the spacing to suit each line length
- keeping the line spacing constant across all line lengths

Question: If you were asked to advise a researcher who was interested in finding the optimum line length for reading from screen, which of the two options above would you recommend? Why?

Comparing typefaces

An even greater problem arises when more than one type of variation is built into the test material. The classic example is the comparison of a serif and a sans serif typeface. If a difference in reading speed is found this could be due to the presence or absence of serifs but also could be due to other ways in which the two typefaces differ (e.g. contrast between thick and thin strokes). Researchers may be insensitive to the confounding variables (that also change along with the variable of interest) but their existence may invalidate the inferences that can be drawn. If we are less concerned about which stylistic feature of the typeface contributes to legibility and more interested in the overall effect, the results may be valid.

Numerous studies have compared the legibility of different typefaces[2] despite potential difficulties in deciding how to make valid comparisons. As a typeface has various stylistic characteristics, which have been shown to affect legibility, comparisons need to consider:

- How to equate for size. Although this may seem straightforward to many people, those with typographical knowledge are aware that typefaces appear to be different sizes depending on the height of the ascenders and capitals, the x-height, and the size of the counters (space within letters). Making sure that the typefaces are matched for their x-height, not point size, helps to make them appear similar in size[3] (see Figure 4.5).
- How to control for differences in weight and width, stroke contrast, and serifs.

Figure 4.5: The pair on the left compare 24 point Georgia with 24 point Garamond; Georgia appears to be quite a lot larger. To make both appear a similar size, Garamond needs to be increased to about 29 point (pair on the right).

Collaborations across disciplines have resulted in experimental modifications of typefaces by type designers (Box 4.1). This approach would appear to provide the ideal solution, but requires a significant contribution from type designers.

Box 4.1: Experimental modifications of typefaces

Morris, Aquilante, Yager, and Bigelow (2002) compared a serif and sans serif version of Lucida (Figure 4.6), designed by Bigelow and Holmes

> *...the designers produced a seriffed and sans-serif pair whose underlying forms are identical in stem weights, character widths, character spacing and fitting, and modulation of thick to thin. The only difference is the presence or absence of serifs, and the slight increase of black area in the seriffed variant. (p245)*

Figure 4.6: Lucida Bright and Lucida Sans.

Beier has designed various typefaces specifically for testing (Beier and Larson, 2010, 2013; Beier and Dyson, 2014; Dyson and Beier, 2016). Figure 4.7 shows the fonts used in Dyson and Beier (2016).

Figure 4.7: The fonts designed by Beier which control the variation by adding stylistic features to the first font (top): italic, weight, contrast, and width.

Illustrating test material

Graphic designers work with visual material and can be frustrated to find that many studies reported in journals do not illustrate the material used in the studies. Consequently, we are left to figure out what was presented to the participants. This may reflect the researchers giving priority to the results of the study (illustrating data in graphs). However, some printed journals have imposed constraints, due to economic considerations. Many journals now publish online and include interactive versions of articles, which allow for additional supporting material. This has resulted in the inclusion of more illustrations and greater transparency in reporting the methods, materials and procedures used in the study.

Familiarity

Chapter 1 introduced the view, held by some, that legibility results reflect our familiarity with the test material. According to this view, we will find it easier to read something which we have been accustomed to reading. This seems to make a lot of sense as we do improve with practice. However, this also creates a significant challenge for experimenters. How can we test a newly designed typeface against existing typefaces, or propose an unusual layout, without disadvantaging the novel material? More fundamentally, when legibility research confirms existing practices, based on traditional craft knowledge, can we be sure that these practices are optimal? Might they instead be the forms which we are most used to reading? This conundrum was raised by Dirk Wendt in writing about the criteria for judging legibility (Wendt, 1970, p43).

Some research by Beier and Larson (2013), described more fully in Chapter 6, examines familiarity directly, rather than as a confounding variable which causes problems. This research aims to address how we might improve on existing designs, and not be constrained by what we have read in the past.

Methods

The tools used to measure legibility have understandably changed over time, primarily from mechanical to computer-controlled devices. The older methods are summarised in Spencer (1968) and described in more detail in Tinker (1963, 1965) and Zachrisson (1965). Despite the changes in technology, many of the underlying principles have remained the same, but we now use different ways to capture the data. There are two broad categories of methods:

- objective, measuring behaviour or physical responses
- subjective, asking readers for opinions

Threshold and related measures

As described in Chapter 1, when reading we first need to be able to experience the sensation of images (letters) on our retina. We also know that we read by identifying letters which we then combine into words (Chapter 2). With this knowledge, it makes sense to measure how easy it is to identify letters or words and we can vary the typographic form (e.g. different typefaces or sizes). One technique used is the threshold method, which aims to measure the first point at which we can detect and identify the letter or word. This might be the greatest distance away or the smallest contrast, or the smallest size of type.

Eye tests are typically carried out in a similar way, obtaining a distance threshold measurement. When having our eyes tested, we may be asked to read from a Snellen chart [4] where the letters decrease in size as we go down the chart (Figure 4.8). We stop at the point when we can no longer decipher the letter and we have reached our threshold. This is letter acuity as the test uses unrelated letters and unconstrained viewing time.

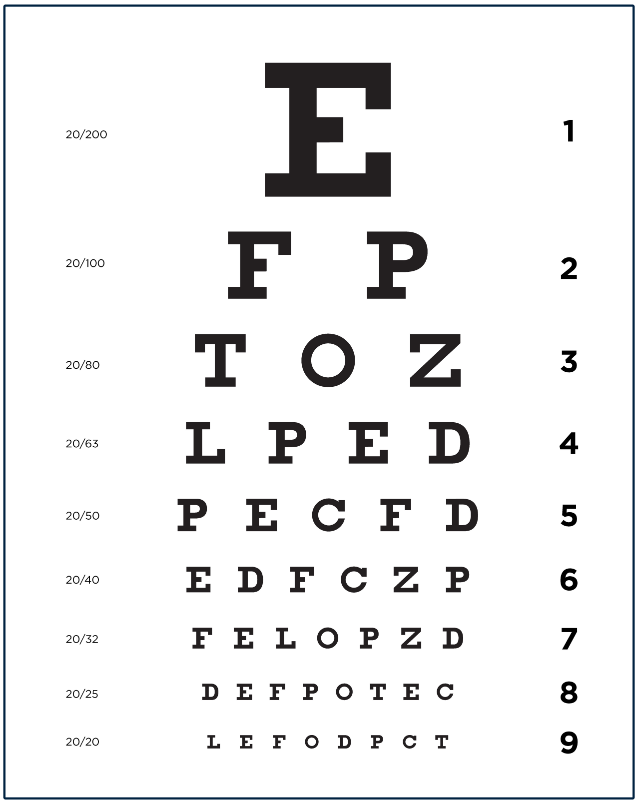


Figure 4.8: An example of the Snellen eye chart, named after a Dutch ophthalmologist in 1862. The smallest letters that can be read accurately indicate the visual acuity of that eye (each eye is tested separately). The bottom row (9) corresponds to 20/20 vision meaning the letters can be read at a distance of 20 feet (about 6 metres).

The eye test uses a similar principle to distance thresholds except the size of type is varied, and we remain seated in our chair at the same distance from the chart. The visual angle is changed in both cases as the visual angle depends on size and distance (see Figure 3.2). In the eye test procedure the visual angle decreases until we can no longer read the letters; distance threshold measures work in the opposite direction with increases in visual angle until we are able to identity the image.

?

Question: Explain why the distance threshold measure needs to start with an image that is too far away to identify and is then moved closer. If you are not sure, read on to find the answer.

The accounts of older methods to test legibility include descriptions of tools which measured thresholds and more general approaches to using thresholds:

- The visibility meter used filters to vary the contrast between the image and the background. The aim was to identify the smallest contrast that still preserves legibility. This has been used to measure the relative legibility of different typefaces using letters or words.
- The focal variator used a similar principle to the visibility meter with a blurred image projected onto a ground glass screen and a measurement was made of the distance at which the image becomes recognisable. This device was limited to using letters.
- A more general method of measuring distance thresholds, which is still in use, is simply to find out how far away something can still be recognised by starting at a great distance and gradually moving the material closer to the participant. The answer to the question above is that it is necessary to do the test in this direction as we cannot accurately report when we can no longer see something because we have already identified it. The method is appropriate for testing signs or other material that would normally be read at a distance but is also applied in other contexts. (See Chapter 5)

- A similar principle is applied when measuring how far out into the periphery an object (e.g. letter) can be placed and still be recognised. Participants are asked to fixate on a specific point, so that they do not move their eyes to focus on the object. Our visual acuity for letters in peripheral vision decreases with eccentricity (i.e. distance from the fovea).

Panel 4.3 describes a sophisticated means of using the threshold to take account of differences among readers.

Panel 4.3: Setting a level of difficulty for each person

The threshold approach can also be applied in a more flexible manner to control how easy it is for a participant in a study to identify letters or words, to improve the sensitivity of the measure. The technique adjusts the presentation for each person, either varying the viewing distance or the length of time shown. Rather than just measuring the threshold, this measure is used to ensure that the level of difficulty is set at a certain level above threshold so that the participants in the study do not get 100% correct or close to 0%. For example, if the task of identifying letters is too easy, any effects of typographic form will not be apparent as even if letters are slightly harder to identify, they will still be identified. Similarly, if the task is too difficult, we either cannot provide answers or guess and get most answers wrong. If we can set the difficulty so that some letters can be identified and some cannot, this should help in revealing differences.

People vary, not only in terms of the more obvious characteristics such as eyesight (visual acuity) and reading ability, but also attention, motivation, fatigue, confidence, and anxiety when taking part in an experiment. Consequently it is useful to be able to set a level for each person. This technique may be particularly valuable in relation to inclusive design as it enables testing of participants with a larger range of abilities than some other techniques because the level of difficulty can be adjusted for each participant. The disadvantage of this approach is that additional time needs to be spent before the main experiment can start.

The short exposure method can be used to measure the threshold (how long is needed to identify a letter or word) or to set a suitable level of difficulty for participants. Before computers were routinely used in experiments, a tachistoscope controlled fixation time by presenting and then removing the image. This is now typically computer-controlled and an example of one form of short exposure presentation is Rapid Serial Visual Presentation (RSVP). Single words are displayed sequentially on screen in the same position which means we don't need to make eye movements (saccades).

RSVP has been in used in reading research from 1970, but has recently emerged as a practical technique for reading from small screens as the sequential presentation takes up less space. RSVP has also been developed into apps promoted as a technique for increasing reading speed. The value of RSVP as a research method for testing legibility is that the experimenter can adjust the rate of presenting a series of words, which can form sentences. However, as with some of the other techniques above, it is only possible to investigate typographic variables at the letter and word level (e.g. typeface, type variant, type size, letter spacing).

The above methods related to threshold measures typically ask the participant to identify what they see (e.g. a letter or word). These responses either comprise the results (e.g. number of correct responses) or the distance/exposure time/eccentricity is recorded which corresponds to a certain level of correct answers.

Speed and accuracy measures

As mentioned in Chapter 3 and earlier in this chapter, speed of reading is a common way of measuring ease of reading, even though the primary concern of designers may not be to facilitate faster reading. If the letters are difficult to identify, we make more eye fixations (pauses) and pause for longer, which slows down reading; more effort is also likely to be expended.

Measures of speed are often combined with some measure of accuracy. This might be accuracy of:

- identifying isolated letters or words

- reading words in sentences and continuous text
- proofreading
- remembering (often referred to as recall)
- understanding (comprehension)

Accuracy can therefore go beyond getting the letters or words correct to measures of recall or comprehension. If letter or word recognition is tested, accuracy may be measured together with exposure time. As we can substitute speed for accuracy when we read, some researchers combine these two measures. If I decide to read very quickly, I am likely to remember and understand less of the text because I am trading off speed and accuracy. If continuous text is read, a test of comprehension is important to check that a certain level of understanding is obtained.

Question: Do you think recall or understanding is more important than speed of reading? Are there any circumstances when speed might be more important?

Measuring legibility by the speed of reading continuous text can be similar to the more usual reading situation. Both silent reading and reading aloud have been used by researchers, though silent reading tends to be more common. If reading aloud, the number of words correctly identified can be measured. Comprehension measures for silent reading include:

- summaries of what has been read
- identifying an error in a sentence, which affects the meaning
- cloze procedure where words are omitted at regular intervals within a text and a suitable word must be inserted into the gap
- open-ended or short answer questions
- multiple choice questions

As a researcher, I have made decisions as to which comprehension measure to use. In doing so, I have weighed up the difficulty of preparing the test material with the difficulty of scoring the results. Table 4.1 summarises my assessment of each of the measures in terms of these two considerations. Panel 4.4 explains the reasons for my assessment and some pointers to good practice when carrying out a study.

Table 4.1: what to consider when choosing a m comprehension

	EASY TO PREPARE	REASONABLY EASY TO PREPARE	QUITE DIFFICULT TO
Easy to score		Identifying errors	
Reasonably easy to score	Cloze procedure	Open-ended questions	Short-answer questio
Difficult to score	Summaries		

Panel 4.4: Considerations when planning comprehension tests

- Summaries require no preparation of questions but the accuracy and completeness of the responses are the most difficult to assess. Decisions need to be made as to whether responses are 100% correct, or partially correct. This difficulty reduces the reliability of the scores.
- This is true to a lesser extent with open-ended questions, as the responses will be more focused and constrained and therefore a little easier to score.
- The cloze procedure is similar to summaries in terms of preparation as it is straightforward to delete words but the responses require judgements as to what are acceptable synonyms as the precise word will not always be inserted.
- Short-answer questions can be more targeted, removing some ambiguity from the assessment.
- Multiple-choice questions are straightforward to assess.
- There is a trend towards the easier the responses are to score, the more difficult the preparation. The exception is identifying an error in a sentence which has the advantage of being relatively easy to prepare and score.

Why are specific questions difficult to create? As with all measures, these questions need to be sufficiently sensitive to detect different levels of comprehension. If the texts are factual, you also need to consider whether participants might know the answers before reading the text. This may require a test of prior knowledge, such as a pre-test (before the main study). The score then becomes the difference between the pre- and post-test, the latter taking place after reading the text. The most difficult questions to generate are multiple-choice as the incorrect alternative answers need to be plausible to make the questions sufficiently difficult.

It is good practice to pilot questions that will be used in a study to detect any problems, such as too easy or difficult, ambiguities, misleading or confusing elements. A pilot is a small-scale study, with maybe only 2 or 3 people, and need not include all aspects of the experiment.

When comparing results across different texts, with different content, the questions on each text need to be at a similar level of difficulty and answers located in similar regions of the texts. Likewise, when identifying errors, the particular words changed, their position, and how they are changed requires careful attention. Various standardised tests have been developed which address these issues:

- Nelson-Denny test (1929) is a multiple-choice test.
- Chapman-Cook Speed of Reading test (1923) has 30 items of 30 words each. In each item there is one word that spoils the meaning and the reader is asked to cross out this word. There is a time limit of 1.75 minutes.

?

Question: Which is the word that spoils the meaning in the item below?

If father had known I was going swimming he would have forbidden it. He found out after I returned and made me promise never to skate again without telling him.

- Tinker Speed of Reading test (1955) is similar to Chapman-Cook but with 450 items of 30 words each. The time limit is 30 minutes.

Question: Which is the word that spoils the meaning in the item below?

We wanted very much to get some good pictures of the baby, so in order to take some snapshots at the picnic grounds, we packed the stove into the car.

Some authors refer to speed of reading as 'rate of work'. This more generic term can cover other types of reading such as scanning text for particular words (as you might in a dictionary or if you are looking for a particular paragraph in a printed text), skim reading or filling in a form.

Physiological measures

In the methods described above the measure is the participant's response, or how fast they respond, or some aspect related to the material (e.g. exposure time, distance from material). Another approach is to take physical measurements of the participants which have included pulse rate, reflex (involuntary) blink rate, and eye movements. These have been described as unconscious processes (Pyke, 1926, p30) which are automatic, whereas we are conscious of threshold, speed, and accuracy measures. An increased pulse rate is supposed to indicate that the participant is working harder. Similarly, an increase in blink rate is assumed to mean that legibility is reduced. However, in both cases, other (confounding) factors may be influencing the measure.

Eye movement measurements, also described as eye tracking, have survived as a technique and now use far more sophisticated technology than the original work around the beginning of the twentieth century (see Chapter 3: Historical perspective). The most widely used current technique records movements by shining a beam of invisible light onto the eye which is reflected back to a sensing device. From this, it is possible to calculate where the person is looking. Typical measurements include:

- frequency or number of fixations (pauses)
- duration of fixations
- number of regressions

The advantage of looking at these individual measures, rather than overall reading speed, is that there may be a trade-off between the number of fixations and their duration. We may make lots of fixations, but for a very short time; conversely we may make few longer fixations. Both may result in the same overall reading time. Regressions indicate a difficulty in identifying letters or words, requiring back-tracking to re-fixate on the relevant part of the text. Another advantage of this technique is that we can measure reading of continuous text in a reasonably natural situation. It is not entirely natural as participants commonly need to wear devices strapped to their head. Eye tracking is also used to explore specific regions of interest (ROI) in advertisements or web pages to see what attracts attention.

Although introduced to measure reader's emotions, changes in facial expression may also indicate the degree of effort exerted and therefore ease of reading (Larson, Hazlett, Chaparro and Picard, 2006). Facial electromyography (EMG) measures tiny changes in the electrical activity of muscles. The muscle which controls eye smiling, for example, is thought to be more of an unconscious process and may therefore reflect emotion or effort which might not be reported (see Subjective judgements below).

As mentioned above when describing how we read different typefaces (Chapter 2), electroencephalography (EEG) technology has recently been applied in research looking at letter recognition. Although the objectives of this research were not to investigate legibility issues, differences in the level of neural activity were found for low and high legibility typefaces [5] . This method may therefore have potential as a means of measuring brain activity to infer how typographic variables influence legibility.

Subjective judgements

This procedure asks people what they think of different examples of material in relation to a particular criterion. Visual fatigue has been measured in this way, by asking people to rate their fatigue on a scale from no discomfort to extreme discomfort. Mental or perceived workload has also been assessed using the NASA Task Load Index (NASA-TLX). As these estimates can be influenced by other factors, a more reliable measure is to test visual fatigue objectively (as a physiological measurement). This has been done using equipment which can simultaneously measure pupillary change, focal accommodation, and eye movements.

A common way of employing subjective judgements in a study is to ask participants which material they think is easiest to read, or which they prefer. These judgements are quite often combined with other methods, such as speed and accuracy of reading. The procedure can vary from asking the participant to rank or rate a number of alternatives to asking them to make comparisons of pairs. (Panel 4.5)

Panel 4.5: Different ways of collecting subjective judgements

Ranking {#ranking .heading-in-Panel}

Ranking asks a participant to put a number of examples of material (e.g. 1) in an order where 1 may indicate the easiest to read and 8 the most difficult to read. This method is suitable if there aren't too many examples to rank. It becomes rather difficult to make comparisons of this nature if there are about 10 or more examples.

Rating {#rating .heading-in-Panel}

Rating can be easier than ranking with many examples as the participant gives a rating for each individual sample, rather than comparing all the samples together. Participants may make some comparisons when rating, but these are not a requirement. The rating scale can be various lengths, e.g. from 1 to 5, or 1 to 7, where 1 might indicate 'very easy to read' and 5 (or 7) might indicate 'very hard to read'. This technique differs from ranking, even though the judgement appears very similar, because there is no need to place the examples in an order.

We should realise that participants will vary as to how they use a rating scale. Some people may use all the scale, e.g. from 1 to 7; others may not use the extremes so that the example they think is the easiest to read may be given a 2 or 3, because it is not thought to be 'very easy to read'. For this reason, researchers sometimes encourage participants to use the full scale.

If the scale has a range which is an odd number (i.e. 5 or 7) this allows for a middle neutral rating which is 'neither easy nor difficult to read' or 'OK'. Some researchers prefer to use a rating scale with an even number to avoid a neutral rating, perhaps because it seems like responding 'Don't know'. A middle rating isn't quite the same as 'Don't know'. As long as distinctions are being made between the examples (i.e. given different ratings), the rating scale is serving its purpose. The results are collated for all participants to see whether they agree.

A semantic differential scale is a specific type of scale where adjectives can be used to rate the appropriateness of typefaces for certain purposes (see Figure 4.9). The two ends of the scale (of 5 or 7 points) are labelled with opposite meanings, for example 1 indicating strong and 7 weak; 1 indicating cheap and 7 expensive. A set of scales using quite a lot of different paired adjectives is given to participants and a statistical technique (factor analysis) determines a smaller number of concepts which underpin all the other adjectives ratings. These describe the nature of the typefaces.

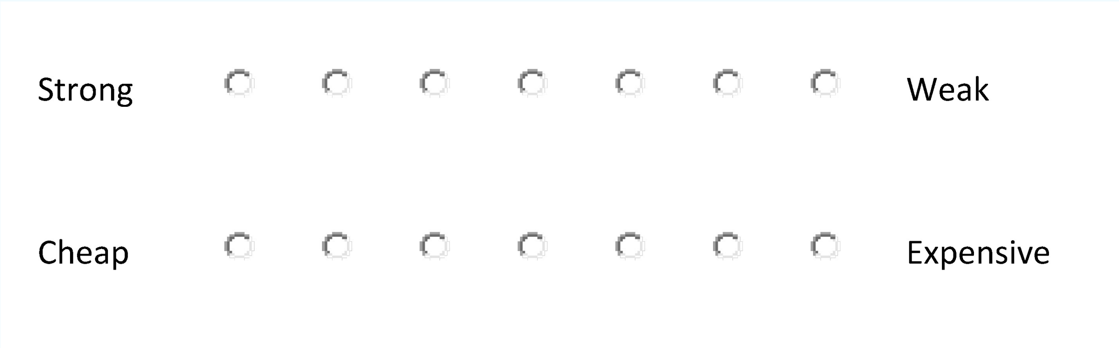


Figure 4.9: Semantic differential scales for two dimensions. The participant is asked to select the circle which best represents their judgement.

Paired comparisons

Another way of making the task of comparing a large number of samples easier for participants is to compare pairs, rather than comparing the whole set at once (ranking). Each sample is compared with every other one, which makes quite a lot of comparisons. However, it is easier to be more confident in saying A is easier to read than B, B is easier to read than C, etc., than putting a large set in a ranked order. This method also detects any uncertainty or inconsistency as if a participant responds:

- A is easier to read than B
- B is easier to read than C
- C is easier to read than A

they are being inconsistent and this might mean that they don't have any strong views about the differences. It may be tempting as an experimenter to include the option of 'Don't know' when using paired comparisons. I advise against this as inconsistencies will reveal this uncertainty without giving participants the ability to opt out with 'Don't know'. As a participant, it may be rather tempting to use 'Don't know' a bit too often. With paired comparisons, as opposed to a rating scale, it is unhelpful to have 'Don't know' responses as they are missing data.

Summary

Having a range of methods to test legibility can be viewed as positive, as they may have different applications, or may be combined within the same study. However, concerns have been raised as to whether studies of single letters or words can tell us anything about everyday reading. It may be tempting to dismiss results from threshold measures of individual characters but we should remember that reading starts with identifying individual characters. If individual characters cannot easily be identified, there is likely to be a problem in reading. Also, it is frequently easier to find differences when using threshold measurements, than when using measures which are closer to the everyday reading process. It is rather pointless to argue for using a method which will probably not be sufficiently sensitive to detect differences in legibility, assuming they exist. Also, it is not feasible to study the complete natural reading experience which will be influenced by numerous variables.

We do, however, need to be aware of the limitations of methods which do not involve reading continuous text. By showing letters or words individually, the reading environment is changed and the effects of many typographic variables cannot be assessed. We are unable to test the effects of changes to word spacing, line length, line spacing, number of columns, alignment, margins, and headings. If we wish to investigate these aspects of typography, we will probably need to more closely approximate natural reading conditions.

The objectives of the study will also guide the choice of method. We should make a clear distinction between testing alternatives as part of the design process and research studies which are intended to inform researchers and designers. In evaluating the value, appropriateness, validity and reliability of any study, the context will determine how and what we measure.

Chapter 5

Overview of research

Introduction

Legibility research up to about the 1980s explored printed material. Herbert Spencer who had written an overview of legibility research in 1968 commented soon after that legibility research needs to be about different forms of output and all media (Spencer, 1970, p73). I doubt he could have envisaged the current ubiquity of mobile phones and tablets, but he did realise that new issues would arise without the constraints of the printed page. Although some relatively recent studies have focused on print legibility (e.g. Lonsdale, 2006, 2007), investigations of reading from and interacting with screens are probably now the more common interests for legibility research [1] .

In providing an overview of the outcomes of legibility research, I am starting with type and building up from there. Research using material presented on screens is discussed together with print, and comparisons made where relevant. Rather than simply summarise the results, I also include the context and objectives of the research because these can affect how we interpret the results and relate them to design practice. Although you may think that a clear set of guidelines and recommendations on how to design to optimise legibility may be more helpful, these would probably oversimplify and mislead. I think a better approach is to try to understand how and why typographic and graphic variables affect different aspects of reading to inform design decisions, rather than prescribe how to design. A set of guidelines based on research are available covering web design and usability.

Screen versus paper

A starting point for research into reading from screen was comparisons with paper; Dillon (1992; 2004, Chapter 3) reviews these studies. In a sense, these were legibility studies as they used measures such as speed of reading and the results usually indicated that reading from screen was slower. At the time, they were helpful in informing educators, but had limited practical application for designers looking for guidance on optimal legibility, unless they were deciding between using screen or print. The results have less relevance today as these older studies from the 1980s and 1990s used cathode ray tube (CRT) technology, now obsolete and replaced with thin film transistor liquid crystal displays (TFT-LCD). These have the advantages of higher display resolution and other improvements in image quality and text presentation capabilities. Panel 5.1 describes a study looking at anti-aliasing and whether this improves legibility.

Panel 5.1: Description of anti-aliasing technique: sub-pixel rendering

Operating systems now use techniques of anti-aliasing and sub-pixel rendering which means that text on screen is close to the quality of printed text. An example of sub-pixel rendering is ClearType, developed by Microsoft in 2000. This technology renders text on screen by separately addressing red, green, and blue sub-pixels with the aim of increasing text legibility.

A relatively small number of studies have tested reader performance and preference with ClearType. The results are somewhat inconsistent which might be due to the different tasks, the choice of technology for comparison, and individual preferences for colour filtering. ClearType text has been found to increase reading speed when compared with non-ClearType (Dillon, Kleinman, Bias, Choi, and Turnbull, 2004; Slattery and Rayner, 2010) but no functional improvements were identified when compared with perceptually-tuned grayscale, a different level of ClearType (Sheedy, Tai, Subbaram, Gowrisankaran, and Hayes, 2008). In this study, moderate ClearType rendering was preferred to text with grayscale or higher-level ClearType contrast, being perceived as improving clarity and contrast.

Along with backlit LCD displays we have dedicated e-book readers with electronic paper or electronic ink (eInk) screens deliberately resembling paper. Given the changes in technology, there are now fewer differences between material in print and on screen and readers also have greater familiarity with reading from screens [2] .

However, some ergonomic differences remain, particularly with desktop computers, such as the distance between reader and material (greater distance for screens), and angle of material to reader (Figure 5.1). Other differences between print and smaller screens (tablets and phones) are primarily related to how text is structured and how we interact with it, and possibly less to do with reading at the level of individual letters and words.

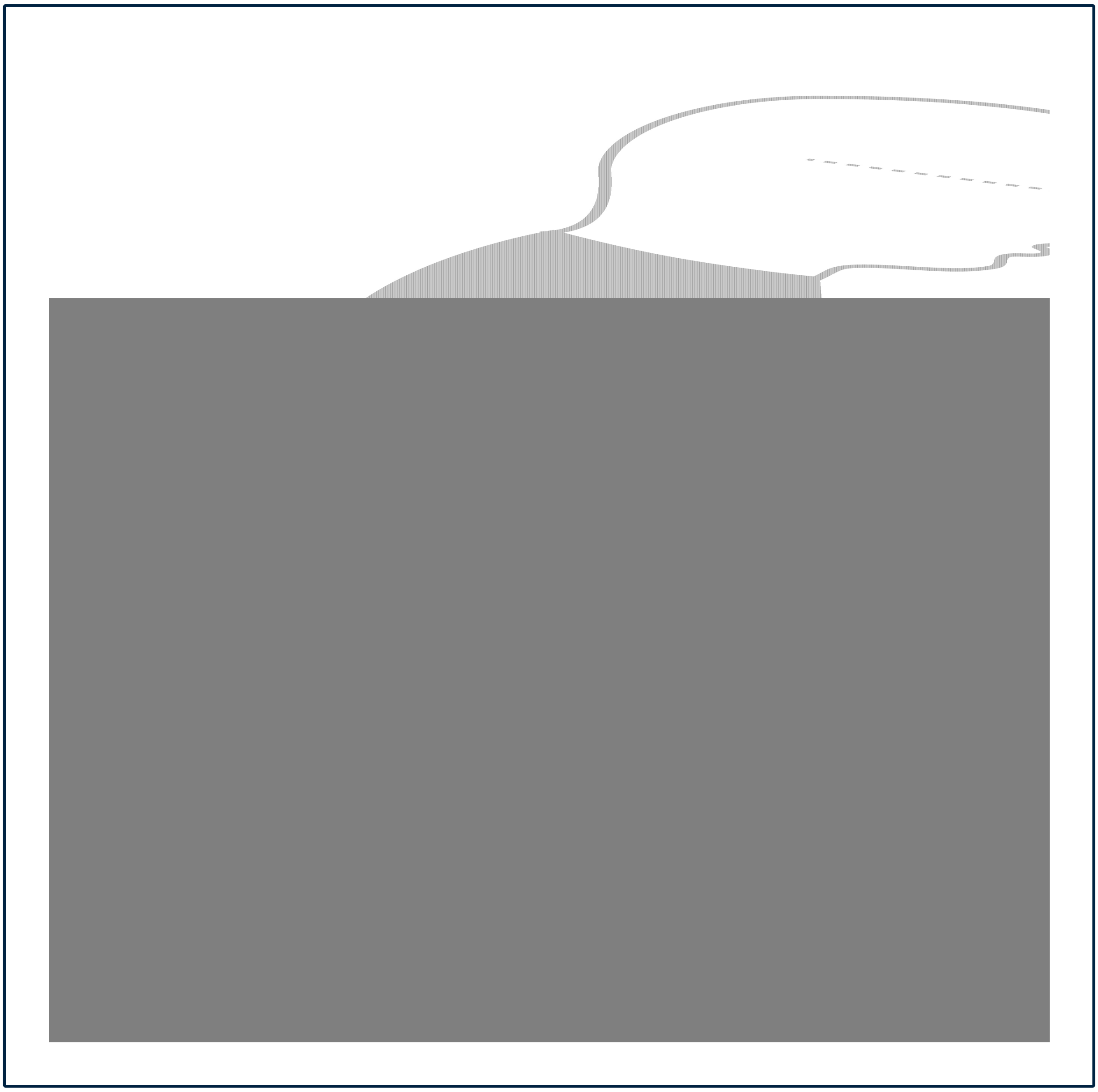


Figure 5.1: The distance between reading material and our eyes and the angle of viewing varies depending on the device. This means that the visual angle of type is relevant as the same type size will subtend a smaller angle at a larger distance (see Figure 3.2).

There has been a recent revival of studies comparing reading from screen and print. These have sought to discover whether reading from screen is still more difficult than reading print. The results suggest that the legibility of text on screen is no longer a problem, although positioning the screen to resemble the angle at which paper is normally read (a display inclination angle of 15°) may be necessary to reduce eyestrain.

Box 5.1 describes more details of the study.

Questions: Do you prefer reading from a screen or paper? Does this depend on what you are reading? Might your preference change if you used a non-preferred method for a reasonably long period of time? Do we simply prefer what we are most used to doing?

Box 5.1: Details of recent study comparing paper and screen reading

A recent study by German psychologists (Köpper, Mayr, and Buchner, 2016) comparing paper and screen used:

- an Apple MacBook Pro with a TFT-LCD widescreen display, backlit by an LED, on a 15.4 inch screen at a resolution of 1680 x 1050 pixels, 128 ppi (pixels per inch)
- an Apple iPad 2
- a 210 x 297 mm sheet of white high quality paper printed using a 600 dpi (dots per inch) laser printer

They measured proofreading accuracy and speed of reading and found no differences between screen and paper. However, screen reading resulted in reports of a stronger degree of eyestrain and reading print was preferred. Reducing the screen luminance did not help but using an iPad which was displayed at the same angle as paper removed the difference in eyestrain symptoms and increased proofreading speed from screen.

Type

A perhaps surprising conclusion from various studies is that typefaces in common use for text (as opposed to display or ornamental typefaces) do not show differences in performance, typically measured by speed of reading and comprehension (Figure 5.2). The traditional research studies are summarised in Tinker (1963, 1965).

Figure 5.2: One of the traditional studies included seven frequently used typefaces and three radically different ones. Cloister Black was read the slowest; Garamond was one of the seven which showed no differences in reading speed (Paterson and Tinker, 1932 summarised in Tinker, 1963, 46–47).

Comparing on-screen typefaces, even those specifically designed for screen (e.g. Georgia, Verdana, Trebuchet, Tahoma) we find that they may not help us read faster, but also do not slow us down. Differences emerge with rather obscure and unusual typefaces that look radically different to the others. For example, a difference is found when comparing Tahoma (sans serif) with an ornate typeface, Corsiva (Figure 5.3).

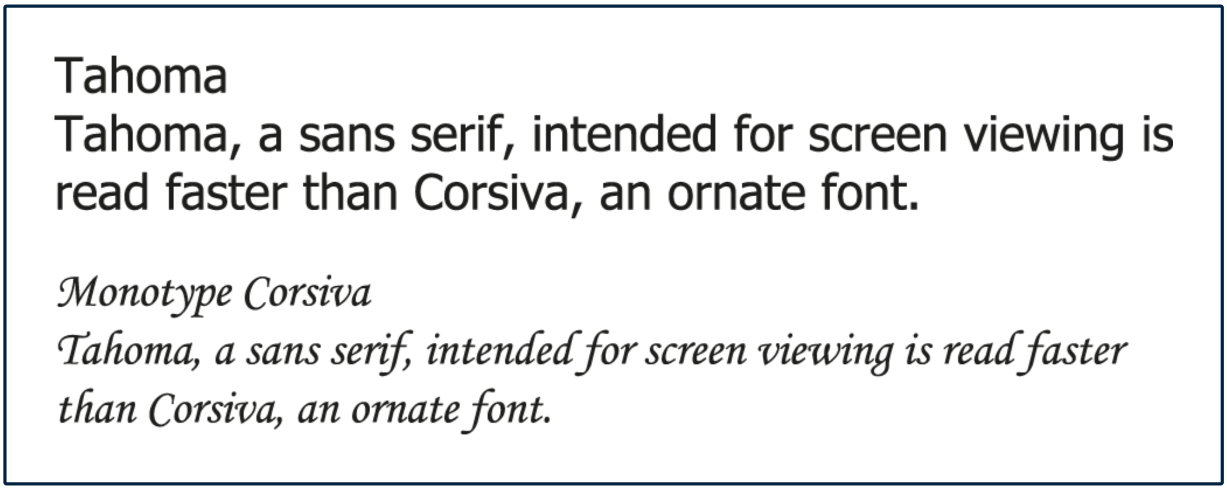


Figure 5.3: As with print-based studies, differences emerge only when comparing text typefaces (e.g. Tahoma) with ornate typefaces (e.g. Corsiva) (Bernard, Mills, Peterson, and Storrer, 2001).

Readers' opinions of relative legibility (subjective judgements) do discriminate between typefaces but this is not usually linked with differences in how they are read (Box 5.2). On the whole, typefaces which have been designed for screen, or are used frequently, are perceived as easier to read and preferred (Boyarski, Neuwirth, Forizzi, and Regli, 1998; Bernard, Mills, Peterson, and Storrer, 2001). They tend to have a larger x-height, wider characters, more open counters and less variation in stroke width.

Box 5.2: Details of study comparing perceived and actual legibility

A study carried out as part of an undergraduate dissertation at the University of Reading, UK looked at the link between how readers judge legibility and how well they identify words (Thompson, 2009). It also explored whether perceptions of legibility change after doing a legibility test. In other words, do readers know how they perform in a test? Are readers able to use their performance to inform their judgements?

Ten typefaces were used (see Figure 5.4) with five described as conventional (Caslon, Courier, Georgia, Helvetica, Times) and five as unconventional (Comic Sans, Corsiva, Curlz MT, Impact, Trajan). These were matched, as far as possible, on the size of the x-heights, not point size (see Chapter 4: Comparing typefaces).

Figure 5.4: Ten typefaces used by Thompson (2009) in his study comparing perceived and actual legibility (subjective and objective measures).

?

Questions: Do you think the categorisations of conventional and unconventional are appropriate? What about Comic Sans? What features or characteristics make a typeface conventional? Which category of typeface (conventional or unconventional) do you think is more legible?

Two groups of participants were tested: ten designers and ten non-designers. The study proceeded as follows:

- Compare pairs of typefaces (see Chapter 4: Paired comparisons) and identify the one perceived as more legible
- Report single words presented on screen for a short time

- Again compare pairs of typefaces and identify the one perceived as more legible

As you probably can predict, the conventional typefaces were judged as more legible than the unconventional. This was true for designers and non-designers although the difference was more marked for designers. The pattern of results was essentially the same before and after the word identification test; Caslon and Georgia did switch places but their scores were very similar. (Figure 5.5)

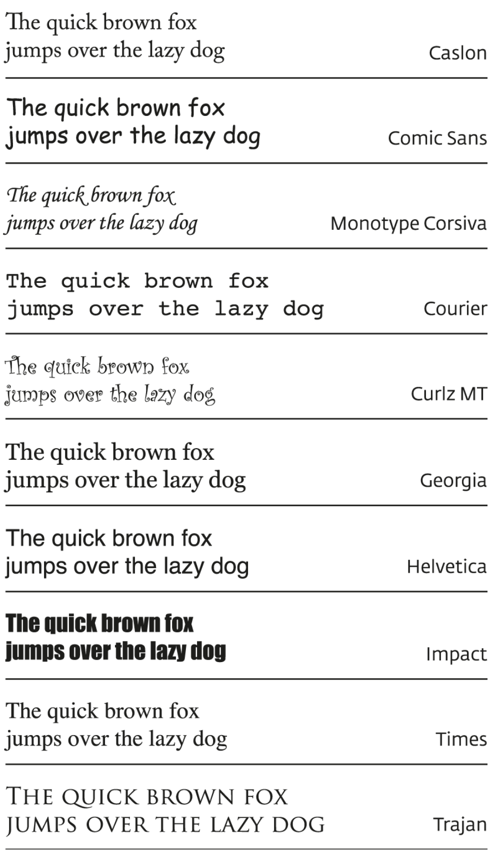


Figure 5.5: Ranked order from most to least legible based on paired comparisons before word recognition task.

?

Question: Can you suggest why designers might have a stronger bias towards perceiving conventional typefaces as more legible than unconventional ones?

The typefaces that were read most easily also grouped according to conventional and unconventional, with conventional more legible. There was one clear exception which was Comic Sans which turns out to be easy to read (Figure 5.6). Yet readers don’t judge it as easy to read (ranked 6 out of 10). Why not?

Rank	Typeface
1	Comic
2	Courier
3	Helvetica
4	Georgia
5	Times
6	Caslon
7	TRAJAN
8	Corsiva
9	Impact
10	Curlz

Figure 5.6: Data from performance test showing the most correct identifications (Comic Sans) down to the least (Curlz).

Occasionally research finds a difference among typefaces when care has been taken to make the experiment as sensitive as possible (see Chapter 4: Challenges). For example, an advantage has been demonstrated for the sans serif Gill Medium over other sans serif typefaces but no differences between serif and sans serif typefaces (Poulton, 1965). His method was to limit reading time of passages of text to 90 seconds and measure how much was comprehended through open-ended questions with short answers (easier to score). This study used two versions of Univers: one matching the other typefaces in terms of x-height; the other matching point size (Figure 5.7).

Serif versus sans serif

One of the more common and somewhat controversial debates concerns the relative legibility of serif and sans serif typefaces. Comparisons of serif and sans serif typefaces typically find no differences in speed of reading or comprehension. In a critical review of 72 studies that compare different typefaces, Lund (1999) found no valid conclusion in favour of either serif or sans serif typefaces. Box 5.3 describes one study supposedly showing an advantage for a serif face.

Question: Why might comparisons of serif and sans serif typefaces be a popular topic for a study?

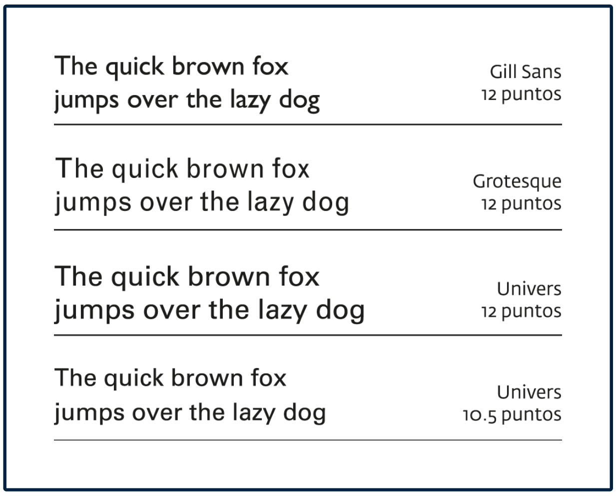


Figure 5.7: The sans serif typefaces used by Poulton (1965) showing the two versions of Univers.

Box 5.3: Critique of study comparing serif and sans serif type

One study which appears to contradict the lack of any reliable difference between reading serif and sans serif typefaces is reported in a booklet and was subsequently incorporated into a book (Wheildon, 1986, 1995). Comprehension was measured for an article with a serif type (Corona) and compared with a sans serif (Helvetica). The results show an unbelievable difference in comprehension:

- 67% of readers had good comprehension levels for serif type
- 12% of readers had good comprehension levels for sans serif type

The size of the difference between these two typefaces is astonishing in comparison with the results of other researchers. Assuming the results are reported accurately, the method of testing may be responsible for the extraordinary nature of the findings. The method is described only briefly, not reaching the standard required for scientific publications, and it is worth noting that this research was not published in an academic journal. Some aspects of the brief account of the method reveal a lack of understanding of experimental procedures. Readers are asked 'leading questions about their attitudes to the articles and layout of the pages' (Wheildon, 1995, p9). Also worrying is Wheildon's concern that the results may have been biased or distorted if he had not done all the work himself.

I include this example because it has been treated seriously by some writers who have not questioned the reliability or validity of the findings. Rather than uncritically citing these results as evidence for differences in legibility, they should be evaluated alongside the majority of other research which has not found the same huge differences.

These comparisons of sans and serif typefaces used existing typefaces which therefore vary in a number of ways other than presence or absence of serifs (Chapter 4: Comparing typefaces). These differences include thickness of stems, lengths of ascenders and descenders, character widths, ratios of thin to thick stroke widths. More recently some studies have aimed to isolate the effect of serifs from these other variables; researchers have found it easier to manipulate typefaces and change individual characteristics with the introduction of digital type. However, expertise is required in these manipulations as there is an interrelationship of elements in a well-designed typeface, within and among letters, which can be disrupted.

This expertise was incorporated into a study carried out by a mathematician, Robert A. Morris, with vision scientist colleagues, by involving a type designer, Charles Bigelow. This study has been referred to above in relation to combining disciplines (Chapter 3) and the challenge of comparing typefaces (Chapter 4). The researchers compared a serif and sans serif version of Lucida, designed by Bigelow and Holmes. The underlying forms are identical with the major variation the presence or absence of serifs which results in a slight increase in the black area of the serif version. They used a small (about 4 point) and large (about 16 point) size and found that serifs slowed down reading at the small size, but there were no differences at the large size.

The sans and serif versions of Lucida have been tested more recently looking at words and sentences [3] .

- With words, the sans serif version was responded to quicker than the serif version (Moret-Tatay and Perea, 2011).
- The second study (Perea, 2013) wished to find out if there is an advantage for a serif typeface over a sans serif during ordinary reading. Publication norms, such as the American Psychological Association (APA), specify that manuscripts should be submitted to journals using a serif typeface like Times New Roman. This might suggest that they believe the text will be easier to read in a serif typeface. The study found that the differences are minimal and did not show the same slight advantage for sans serif found with individual words. (See Box 5.4 for further details of methods of all the Lucida studies)

Box 5.4: Details of studies using Lucida

The study by Morris, Aquilante, Yager, and Bigelow (2002) used sentences presented on screen using RSVP (see Chapter 4: Threshold and related measures) and displayed these at a distance. Characters with an x-height of 40 pixels at a 4-metre distance equates to about 4 point type at a normal reading distance (40 cm). By increasing the size of the type, characters could be rendered appropriately (i.e. sufficient pixels). Displaying a 4 point type on screen might have resulted in problems. By viewing at a distance, the visual angle is reduced and the characters appear smaller. (See description of visual angles, Panel 3.1, as a reminder of the relationship between size and distance). The large (16 point) type was produced with an x-height of 160 pixels at a 4-metre distance.

Moret-Tatay and Perea (2011) used individual words and a lexical decision task which involves deciding whether the item is a word or a non-word. This task requires us to not only identify letters but process them to the point of matching them with a word (or not).

Perea (2013) decided it was important to use a setting closer to typical reading than the lexical decision task and RSVP (used by Morris *et al.*, 2002). One-line sentences in 14 point Lucida or Lucida Sans were read on screen and eye movements were monitored.

Reasons proposed for the advantage of serif typefaces for reading continuous text are that the serifs:

- contribute to the individuality of letters (yes, possibly)
- make words and lines hang together (no)
- guide the eye along the line of text (no)

The first reason is plausible as we know that the individuality of letters is important; the easier it is to differentiate letters, the easier it will be to read. However, there are other means of making letters more discriminable than adding serifs (see Letter features, below). See Panel 5.2 for explanations as to why the last two reasons don't fit with what we know about reading.

Panel 5.2: Critique of the role of serifs in reading

There is no evidence that serifs have the functions of keeping letters in words together or words in lines. These are two quite distinct functions and neither fit with what we know about reading. It is possible that this explanation stems, in part, from the mistaken belief that we use word shape, rather than individual letters, to recognise words. Letter and word spacing (covered below) may affect the ease with which we recognise letters and words.

There are reasons why serifs are unlikely to guide the eye along the line of text. We use our peripheral (parafoveal) vision to guide where we land our eyes following a saccade. The targets for saccades are probably determined by the location of word boundaries. The serifs would be much less effective at guiding the landing points of the saccades because the detail of serifs is largely missing in our peripheral vision; our visual acuity decreases with distance from the fovea.

Individual letters

Tinker explored the relative legibility of lower case letters (summarised in Tinker, 1963). He concluded that some letters are intrinsically more legible than others because they are more discriminable, i.e. they have certain distinguishing features.

- High legibility: `d m p q w`
- Medium legibility: `j r v x y`
- Low legibility: `c e i n l`

Tinker came up with this order from most to least legible:
`k d q b p m w f h j y r t x v z c o a u g e i n s l`

The reasons given by Tinker for these differences in legibility are:

- some letter pairs may be confused such as `c` and `e`; `i` and `j`; `i` and `l`
- narrower letters (`e` or `i`) are less legible than wider letters (`m` and `w`)
- simpler outlines (`w` and `q`) are more legible than more complex outlines (`a` and `g`)
- having a distinguishing characteristic aids legibility such that `b d p q k` will be more legible than `n` and `u`

Another possible reason for differences in legibility isletter frequency. The accuracy of identifying a letter (Larson and Carter, 2016) and the speed of determining whether an item is a letter or a non-letter (New and Grainger, 2011) has been found to correlate with the letter's frequency. We might expect that the more often we encounter a letter, the easier it is to identify. However, not all studies have found this effect, and this includes Tinker who reported no relation or a small negative correlation between letter frequency and legibility with lower case letters (Tinker, 1928).

As we cannot choose to compose a text that avoids letters of low legibility, or low frequency, these deductions are not particularly helpful. They may guide type designers as to where attention might be focused to improve the legibility of their typefaces, or help the design of logotypes. But as graphic designers choosing a typeface for use in particular circumstances (e.g. low illumination) or for specific groups of readers (e.g. visually impaired, beginner readers), we need to know which features of typefaces, not letters, influence legibility. Tinker did include some direction on the shape of serifs and which letters they are applied to; the ratio between thick and thin stroke widths (modulation); and the size of counters (white space within letters). However, these suggestions were not supported by experiments and were also influenced by the printing processes of the time. We should therefore look to more recent research.

Letter features

Psychological research has shown that we detect simple features of letters, in order to identify the letters. These were previously described as ‘distinctive features’ emphasising their role in providing cues to differentiating the letters. In Chapter 2 I mentioned that researchers who develop models of reading have tended to assume that the font will not affect how letters are identified. But more recent research indicates that if there are more features, we are less efficient at identifying the letters. More complex forms, e.g. ornate typefaces, have more features (Panel 5.3). Therefore, what Tinker deduced, but did not test, appears to be correct although he was referring to different letters rather than different typefaces. Simpler outlines are more legible than complex outlines. The example in Figure 5.8 is exaggerated, as we wouldn’t consider a script typeface (with a complex outline) to be suitable for reading continuous text, but a comparison of these two typefaces illustrates the point.

Figure 5.8: Kunstler Script, 36 point (top) is compared with Arial, 24 point (bottom).

Simpler forms of some letters have been developed for children’s reading based on the view held by many teachers that children will find it easier to read letters that are similar in shape to those they write. The modified letters are described as ‘infant characters’ and the differences are most apparent in the `a` and `g` (see Figure 5.9). A study by Walker and Reynolds (2003) found no differences between typefaces with and without infant characters in terms of errors when children between 5 and 7 years old read aloud. The children were aware of the different forms, with some able to identify which they read and which they used in writing. These results suggest that non-infant characters are not problematic and they do not necessarily help in discriminating among letters.

Gill Sans: note shapes of a g & y

Gill Sans Infant: note shapes of a g & y

Figure 5.9: Two versions of Gill Sans showing the modified shape of the `a` and `g` in the bottom row. A child in the study by Walker and Reynolds (2003) also noticed the difference between the `y` in the infant typeface.

Panel 5.3: How perimetric complexity is measured

How do we measure complexity? One way is perimetric complexity which compares the perimeter (the inside and outside edge) to the overall area of the letters which is described as the ‘ink’ (the area covered by pixels). The precise formula is the inside and outside perimeter (p) squared, divided by the total ink (a).

> p^2 / a

A convoluted or elaborate form will have a larger perimeter compared to total area and therefore be more complex. Perimetric complexity is not exactly the same as peoples’ subjective ratings of complexity; these tend to reflect the number of turns in the outline.

(Pelli, Burns, Farell, and Moore-Page, 2006).

This century, a number of psychological studies have aimed to identify the particular features that are most important in identifying letters. Unfortunately, they have produced different answers which means that further research is necessary to clarify our knowledge. The reasons for different conclusions may be because of variation in the:

- way letters are divided into components
- method of testing
- typeface used in the test
- case, either upper or lower case, or both

To add to the confusion, the terminology for the different letter parts doesn't always coincide with the more precise descriptions of type designers. Also, there isn't always consistency in terminology across studies. In outlining the results of four of these studies, I will introduce a consistent terminology. Designers talk of 'strokes' rather than lines, reflecting a calligraphic origin, and the components can be described as:

- stroke terminals or endings (rather than terminations) which will differ in a serif typeface; this means that the features underlying letter recognition may depend on the typeface
- stroke junctions (sometimes called intersections or vertices)
- mid segments which can include vertical, horizontal, diagonal or curved strokes

One of the original studies to explore the role of various components of letters used the typeface Courier (Petit and Grainger, 2002). They found that mid segments of letters play a critical role in letter identification (see Figure 5.10).

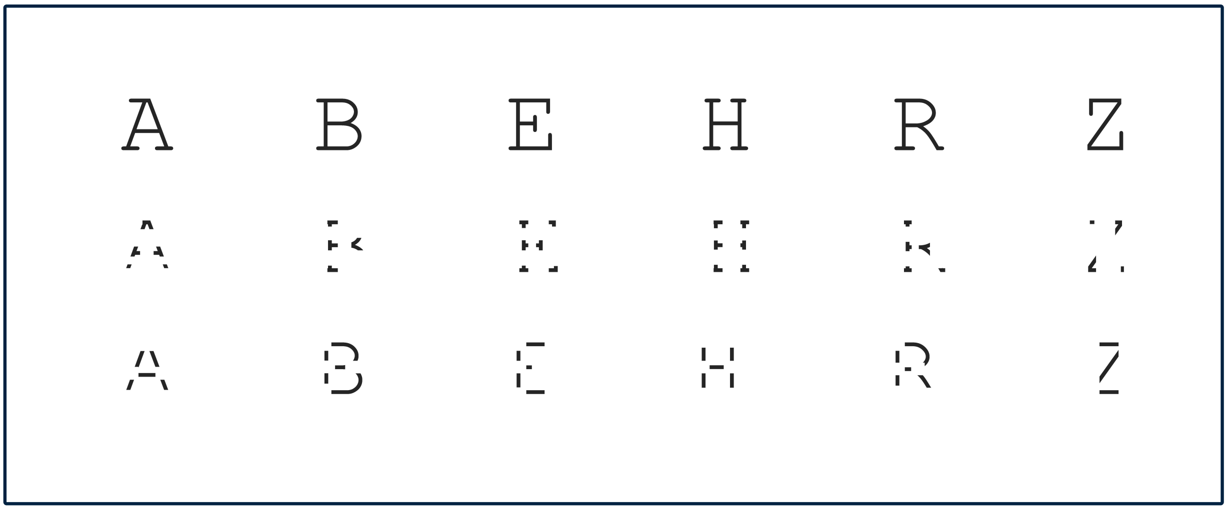


Figure 5.10: Complete letters in Courier upper case (top); letters with only the stroke junctions (middle); letters with only the mid segments (bottom).

Some years later, two studies using the typeface Arial report that stroke terminals (in particular) and horizontal lines (a form of mid segment) are important cues to letter identity for both upper and lower case letters; stroke junctions are quite important for uppercase; and slants tilted right (another type of mid segment) are more useful for identifying lowercase than uppercase (Fiset, Blais, Éthier-Majcher, Arguin, Bub, and Gosselin, 2008; Fiset, Blais, Arguin, Tadros, Éthier-Majcher, Bub, and Gosselin, 2009). Figure 5.11 shows these parts of letters.

Figure 5.11: The parts or features of letters that have been found to be important for recognising the letters and distinguishing them from others. The top two rows show the stroke terminals or endings that are important in letter identification in upper and lowercase. The two middle rows show the horizontal strokes, again used in identifying upper and lower case letters. The last two rows show the stroke junctions that are most relevant in identifying upper case letters, and the diagonal strokes tilted right, used as cues in lower case.

Around the same time, Lanthier, Risko, Stolzh, and Besner (2009) found that taking out the stroke junctions from Arial Narrow upper case letters makes letter and word identification more difficult compared to taking out the mid segments (see Figure 5.12). This suggests that stroke junctions are important in letter identification.

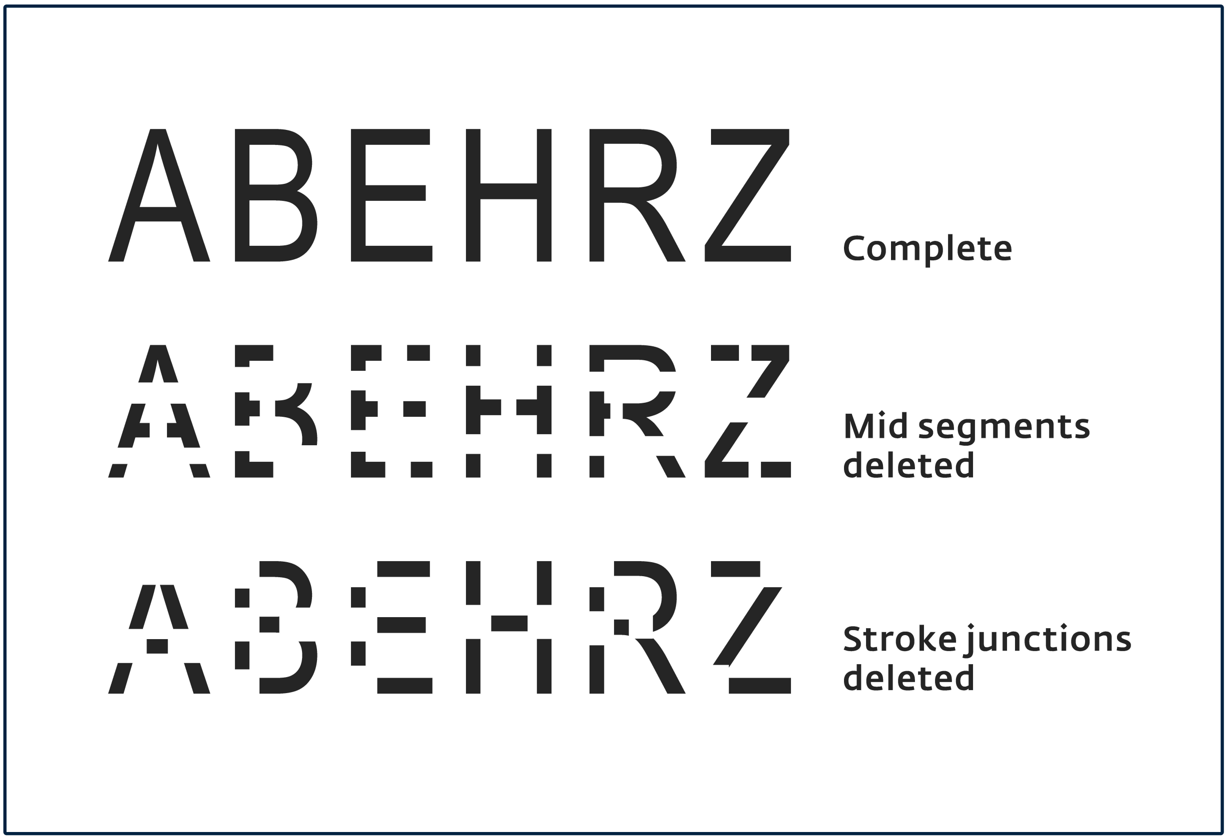


Figure 5.12: Complete letters in Arial Narrow upper case (top); letters without the mid segments (middle); letters without the stroke junctions (bottom).

The fourth study, again exploring which components of letters are more important in words, uses the typeface Minion (Rosa, Perea, and Enneson, 2016). Their results show that the mid-segments are the most important for identification, followed by stroke junctions; terminals do not appear to be critical. Figure 5.13 illustrates the removal of each of the three components as this manipulation was used to determine the contribution of each. As this study used a serif typeface (in contrast to a sans serif, Arial), removing terminals may have less impact.

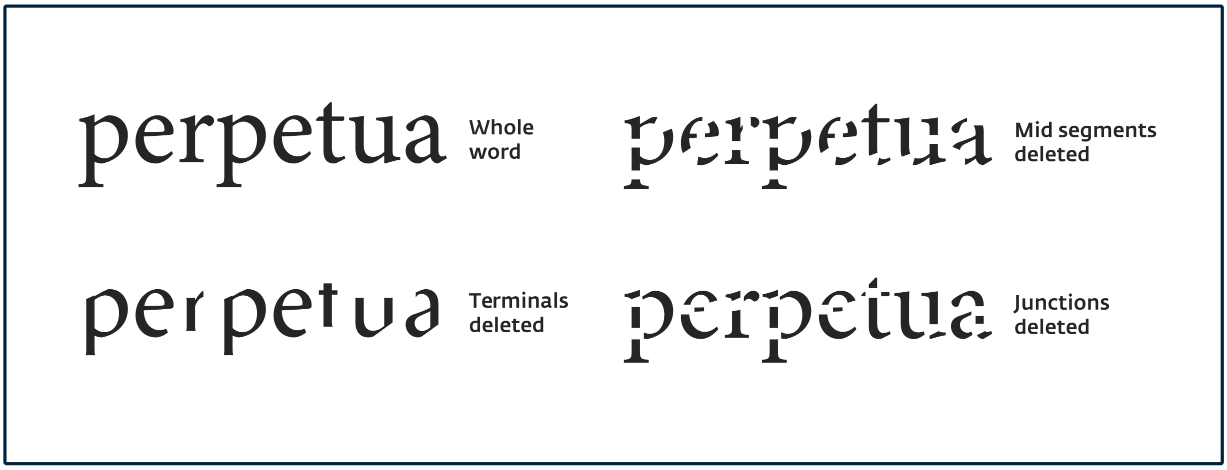


Figure 5.13: Four versions of the word perpetua starting with the whole word and showing the three types of deletions: terminals deleted, junctions deleted, mid-segments deleted (Rosa, Perea, and Enneson, 2016.)

Box 5.5 provides more detail on how these four studies divided letters into components and the different methods of testing.

Box 5.5: Methods used to identify letter features

One way in which the procedures used in the studies vary is whether components of the letters are removed or certain components selected for inclusion. It may seem as though the outcome would be the same but this is not the case as there are other parts to the letter (see Figure 5.14). The fourth study I reported (Rosa, Perea, and Enneson, 2016) includes both procedures: they started by including components and did not find any differences among mid segments, stroke junctions and terminals. When they changed to deleting each of the components, they did find differences (see Figure 5.13).

Figure 5.14: Mid segments and junctions are included in the letters of the word (left); terminals are deleted (right). Based on Figures 2 and 3 of Rosa, Perea and Enneson (2016).

The methods used to measure letter or word identification include:

- priming with alphabetic decision or letter identification: a full letter or part of a letter (Figure 5.10) is shown for a very short time (30 or 50msecs) and then the same complete letter is shown and the participant says whether it is a letter or not (alphabetic decision) or says which letter it is (letter identification)
- delayed segment with lexical decision: a part of the word is displayed very briefly followed by the whole word and the participant says if it is a word or not a word
- straightforward letter and word identification, i.e. name the letter or word
- a classification image technique which essentially varies the amount of the letter displayed over time and the participant identifies the letter

Comparing the results from the four groups of researchers indicates that we don’t yet have a clear picture of how we identify a letter. Two studies highlight mid segments as important, one stroke junctions and another terminals. As yet, I am not aware of any study which compares different typefaces (upper and lower case) using one of these methods to see if the components or features we use to differentiate letters depend on the typeface characteristics [4] .

A more typographical perspective is to look at serifs which can function as terminals and may contribute to differentiating some letters. But why do no clear differences emerge when serif and sans serif typefaces are compared? A possible contributory factor is that serifs can improve the discriminability of some letters (i.e. make them less similar to other letters) but serifs may also make other letters less discriminable, and therefore liable to misrecognition. Therefore at some stroke endings serifs may help, but not at all endings. Box 5.6 describes some studies which address this issue focusing on specific letters.

Box 5.6: Details of studies looking at the contribution of serifs

Some time ago, Harris (1973) compared the legibility of individual letters in two sans serif typefaces (Univers 689 and Gill Sans Medium) and one serif typeface (Baskerville 169). The letters were shown off centre in a tachistoscope, for brief viewing. His results suggested that serifs can close up open counters, impairing recognition, but in other letters the serif enhances gaps. As he used existing typefaces, the results may be attributed to aspects of the typeface other than serifs (e.g. x-height, letter contrast, weight).

A study by Beier and Dyson (2014) followed up on this looking at the same individual lower case letters (`j, i, l, b, h, n, u, a`). The letters were set in the typeface Ovink, a sans serif typeface designed for distance viewing, and a new slab serif version which differed only in relation to the added serifs. The typefaces are designed by Sofie Beier (see Figure 5.15).

Ovink sans

Ovink serif

Figure 5.15: The two versions of Ovink differing only in relation to the serifs.

In this case, the role that serifs play when letters are viewed at a distance was explored. We found that serifs at vertical extremes (`l, b, h, n, u`) facilitate letter recognition but in letters `i` and `j`, serifs do not help. The serif is not at the vertical extreme because of the dot. In these letters, the serifs may remove the narrow character of these two letters resulting in lower legibility. See Figure 5.16.

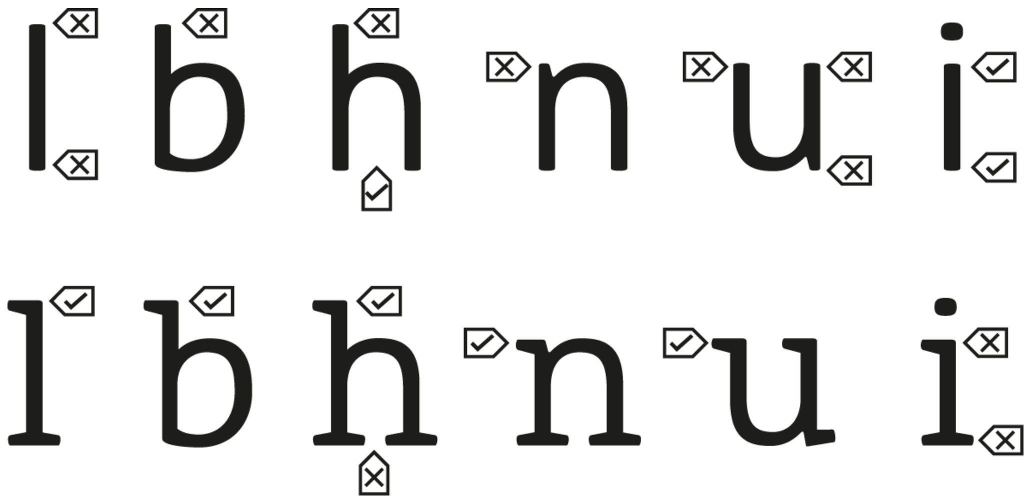


Figure 5.16: Recommendations from Beier and Dyson (2014): remove serifs when not at extremes (`i`); `h` can be confused with b so recommended that serifs removed from the counter of `h`.

These outcomes support the general conclusion that serifs can be both helpful and unhelpful in letter recognition. This tends to make choosing a typeface rather complicated but helps to explain why we don't find differences in performance when comparing serif and sans serif.

Unfortunately, the conclusions are even less straightforward as different results can be obtained when looking at individual letters viewed in parafoveal vision (off centre), rather than at a distance. This has important implications for designers in choosing typefaces for specific contexts.

Another approach to determining which features influence legibility has focused on individual letter confusions such as those proposed by Tinker (i.e. c and e). The objectives were to provide recommendations for specific design elements for onscreen reading and situations where codes or single characters need to be quickly and accurately identified, such as air traffic control displays (Fox, Chaparro, and Merkle, 2007). Box 5.7 describes what this study found.

Box 5.7: Outcomes of study looking at letter e

We do have some insight into the particular difficulties with the letter e, which can be mistaken for a c or o. Comparing 20 typefaces, an e in Verdana was always correctly identified whereas an e in Garamond was only correct 10% of the time. Using a statistical procedure, the researchers determined that the problem with Garamond is the higher bar compared to overall height. Surprisingly, the overall size was not important. Although this result seems plausible, we might instead describe the difference as a smaller counter (see Figure 5.17).

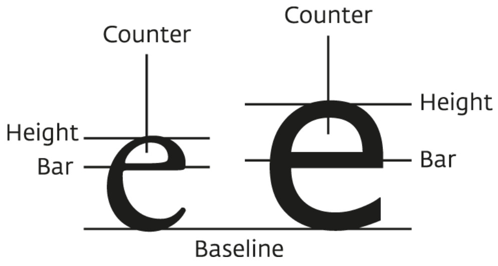


Figure 5.17: Garamond (left) has a higher bar, in relation to overall height, than Verdana (right). This results in Garamond having a smaller counter.

Letter features have been researched from psychological and design perspectives, the former aiming to formulate more general theories of letter processing and the latter focusing on specific details. They therefore complement each other. A useful way forward might be to establish whether the general theories apply to all typefaces by comparing typefaces with very different characteristics.

Upper versus lower halves of letters and words

An effect which can be very easily demonstrated is the relative ease of reading text when only the top halves of letters are available compared with the bottom halves (Figure 5.18). This is obviously not a way in which we would set text, but it may tell us something about how we read, for example through eye movements. This knowledge may help us, perhaps indirectly, in making design decisions.

Huey (1908/1968) observed the advantage to perception of the upper half claiming that

> *…the upper half of a word or letter is obviously more important for perception than is the lower half. Huey (1968, p98)*

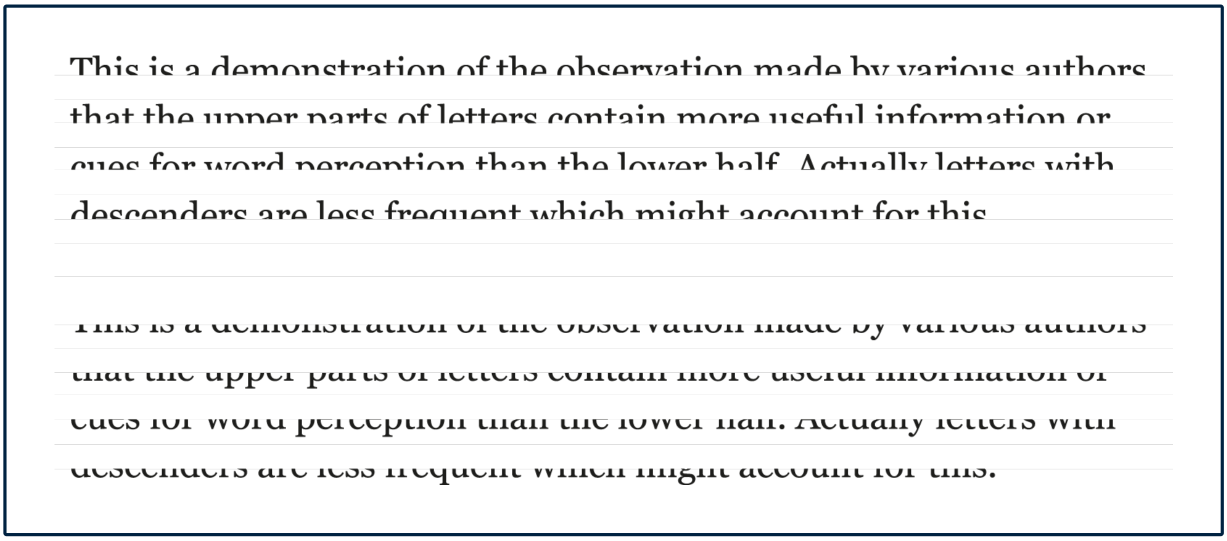


Figure 5.18: It is easier to read the text when the top halves of letters are visible than when we only see the bottom halves.

This was explained in an early printers’ handbook *Typographical printing-surfaces: the technology and mechanism of their production* by Legros and Grant (1916) as more frequent letters projecting above the middle line. Letter frequency counts can vary depending on how they have been counted, what content is used, and the language. However, despite differences the consensus is that the first letter with a descender (`p` or `g` in English and probably `p` in Spanish) is number 16 in terms of frequency; there are 4 or 5 letters with ascenders that are more frequent than `p`. Logically, this tells us that there will be more letter parts above the midline than below which can disambiguate the letter. A comparison of the level of ambiguity in the lower and upper part of letters across some European languages (Tejero, Perea, and Jiménez, 2014) shows similarities:

- English: 68% of letters are ambiguous in lower part; 51% ambiguous in upper part
- Spanish: 68% of letters ambiguous in lower part; 50% ambiguous in upper part
- French: 68% of letters ambiguous in lower part; 50% ambiguous in upper part [5]

By examining eye movements, we know that the eye fixates for longer when reading the lower half compared to the upper half, indicating that removing the upper half produces a greater cost to reading (Perea, 2012). The research described above which identified the features of letters we use to distinguish one letter from another did not find a bias towards features in the upper parts of letters. The bias we see in the demonstration (Figure 5.18) appears to be restricted to letters in the context of words. This is because words do not have equal numbers of each letter but have more letters that are ambiguous in their lower part (in the Latin alphabet). This is evidenced by a clever experiment which controlled the number of ambiguous letters in the top and bottom half of words and removed the effect (Tejero, Perea, and Jiménez, 2014).

Upper versus lower case

Unlike comparisons of different typefaces, a fairly consistent result is that all upper case (capitals) slows down reading compared with lower case or sentence case (where the beginnings of sentences are capitalised). In the past, this was attributed to the loss of word shape (ascenders and descenders) in upper case, but as we read by identifying individual letters, this cannot be the explanation. We are more familiar with reading lower case in continuous text which can account for this advantage [6] . This explanation is proposed by a very recent study which found that when reading sentences, words in upper case were more likely to be re-fixated (looked at again) than words in lowercase (Perea, Rosa, and Marcet, 2017). The researchers suggest that we do an initial familiarity check before we move our eyes to the next location and this check is more likely to be a match with stored words if we are reading more familiar visual forms.

However, at the same point size, upper case is larger than lower case. Should the x-height of lower case be equal to the cap height when we make comparisons in experiments?

- If lower case (bold) letters approximately match the x-heights of upper case (Figure

5.19), headlines are located faster in lower case (Poulton, 1967).

- If we don't adjust but compare Arial in the same point size for caps and lower case (Figure 5.20), upper case appears to be more legible (Arditi and Cho, 2007). This is logical as the upper case letters are larger. For readers with normal vision, reading is quicker for upper case when at acuity limits, but this advantage goes when using a larger size that is typical of regular reading conditions.

Headlines set in bold lower case with the x-height matching the height of all caps

HEADLINES SET IN ALL CAPS WITH THE CAP HEIGHT MATCHING THE X-HEIGHT OF LOWER CASE

Figure 5.19 Comparison of text in Times New Roman 22.5 point bold and Times New Roman 14 point all capitals. The x-height of the upper examples matches the cap height of the lower example by adjusting the nominal point size. With this adjustment, headlines were found faster in bold lower case (Poulton, 1967).

upper case appears to be more legible than lower case if we don't match the x-height

UPPER CASE APPEARS TO BE MORE LEGIBLE THAN LOWER CASE IF WE DON'T MATCH THE X-HEIGHT

Figure 5.20: Lower and upper case both in 12 point Arial.

All of this seems to point to the physical size of letters being important, as well as familiarity, i.e. what we are used to reading.

Question: Are you surprised by this?

Type size

If we remain at the level of letters, explaining legibility would seem to be very straightforward:

> The size and shape of printed symbols determine the legibility of text. Legge and Bigelow (2011, p1)

Shapes have been covered above in some detail and differences between upper and lower case led to the conclusion that size may be more relevant than shape.

One approach to finding out the most appropriate type size for reading continuous text is to determine limits. The smallest character size for which reading is possible at maximum speed is called ‘critical print size’. At sizes smaller than this, reading speed gets much slower. The critical print size depends on individuals, typefaces, and how you measure it. There is also the difficulty discussed above that typefaces of the same point size have different x-heights. Because the smallest or optimal point size for legibility will depend on the typeface, some research will be valid only for the particular typefaces used in the studies.

A way to resolve this issue may have emerged from the collaboration (mentioned previously) between the vision scientist and type designer (Legge and Bigelow, 2011). They take various past studies and translate the type sizes into measurements of the visual angles of the x-heights. To make this accessible to designers, they describe what this would mean in relation to a common typeface. They report that studies indicate that the critical print size is an x-height of 0.2 degrees which is equivalent to 9 point Times New Roman at a distance of 40cm. This happens to be consistent with Tinker finding that 9 point Granjon was read as fast as larger sizes (Tinker, 1963, p71). This convergence of a minimum size for print is encouraging as different methods were used to come to the same conclusion making the result more reliable. However, a distinction should be made between the critical print size (minimum) and the size that optimises reading performance. Box 5.8 gives more details of the collaborative study.

Box 5.8: Details of study by vision scientist and type designer

This collaboration went further in bringing together typography and psychology by considering whether the size of print we use today (and historically) corresponds to the most appropriate size for fluent reading. In other words, have we got it right in the past and present without the specific scientific knowledge that we now have?

The research involves a survey of documents (published books, newspapers, and typefounders' specimens) looking at the size of print and comparing this to what we know about the psychophysics of reading. They found that these sizes fall within the range over which text can be read at maximum speed. They conclude by proposing that the properties of human visual processing play a dominant role in constraining the distribution of print sizes in common use. Their conclusion supports an ecological hypothesis that decisions made by type designers and typographers on type sizes have been determined by properties of our vision.

I suspect that craft experience and practical design skills and training encourage an awareness of the need to attend to perceptions of what we design, not just the objects themselves. A key issue in the study of perception (within psychology) is the potential for the lack of a one-to-one relationship between a physical entity and its perception. This can be best demonstrated with Rubin's vase (Figure 5.21). Do you see a vase or faces?

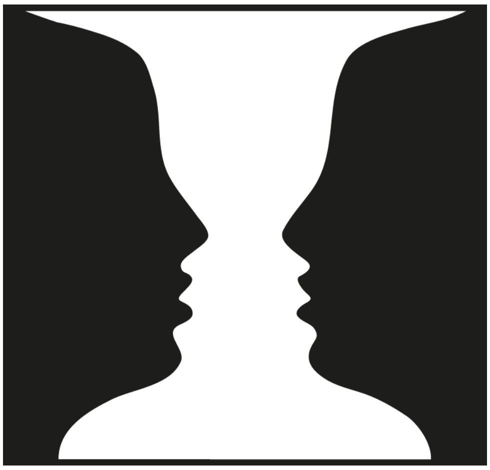


Figure 5.21: Rubin's vase named after the Danish psychologist Edgar Rubin. This is one example of an ambiguous form that has two shape interpretations (perceptions) with only one physical entity and one retinal image. We can only see one perception at a time, but you should be able to switch between the two.

On screen, a slightly larger size of 10 point seems to be required for ‘threshold legibility’, i.e. the smallest size that we can recognise letters and words. The importance of x-height in relation to body size was also found to be a factor in increasing legibility (Sheedy, Subbaram, Zimmerman and Hayes, 2005). For a given body size, Verdana was the most legible and Times New Roman the least legible, with Arial and Georgia intermediate in legibility (see Figure 5.22).

Verdana has an x-height to body size of 0.55

Arial has an x-height to body size of 0.52

Georgia has an x-height to body size of 0.48

Times New Roman has an x-height to body size of 0.45

Figure 5.22: Relationship between x-height and body size (based on figures in Legge and Bigelow, 2011) which correspond to threshold legibility.

When speed of reading is measured, 12 point is read faster than 10 point, but the difference is relatively small. In this study by Bernard, Lida, Riley, Hackler, and Janzen (2002), the researchers found a trade-off between speed and accuracy: the slightly faster reading of 12 point resulted in missing some of the deliberate errors (substituted words) in the text. Some suggestion that there may be no advantage in going above 10 point on screen when using Helvetica and Georgia comes from an eye tracking study (Beymer, Russell, and Orton, 2008). But we must remember that the x-height of the typeface is likely to be the determining factor.

All of the above research relates to adults. Children's reading books typically use larger type sizes and generous line spacing and these both reduce as reading age increases. The key is to ensure that differences among letters are easy to discriminate at early ages so that the child can focus on the other aspects of reading (deriving sound and meaning) rather than perceptual processing, i.e. identifying the letters.

Tinker (1965) proposed that by about 10 years old children respond to typographical arrangements in the same way as adults, therefore at that age between 10 and 12 point type would be suitable. Sizes recommended for younger children are:

- between 14 to 18 point for 5–7 year olds
- between 14 and 16 point for 7–9 year olds
- about 12 point for 9–10 year olds

Question: Based on what you now know about how we read, can you suggest why it may not be a good idea to continue using larger sizes beyond about 10 years old?

Research seems to support the advantage of larger print for younger children and some researchers argue that type sizes in children's reading schemes could be larger than are currently employed (Hughes and Wilkins, 2000).

Type variants (bold and italic)

Traditional research indicates that text set all in italics slows down reading; bold appears not to affect speed of reading continuous text and can be perceived at a greater distance (summarised in Tinker, 1963, 1965). Typical practice and recommendations from well-regarded typographic books such as *The elements of typographic style* (Bringhurst, 1992) is to use bold for setting titles, emphasising keywords etc., and to use italic as a means of differentiating words or sentences within longer paragraphs. These differentiations can be regarded as 'typographic cueing' which can work as an isolation effect, setting apart some information and making it more likely to be noticed by readers.

Looking at how quickly we can recognise a word (by saying whether it is a word or not a word — a non-word), bold words are responded to faster than roman (using Bookman and Arial typefaces). This is particularly the case if the word is uncommon (referred to as low frequency [7]) (Macaya and Perea, 2014).

It may seem that it is a good idea to use a bold font for setting whole texts but some further evidence suggests that a distinction can be made between a font's legibility and the perceptual salience of individual words (Dyson and Beier, 2016). This study explored switching between roman and different variants to see which stylistic features (weight, width, contrast, and italic) disrupt word recognition. We found that single bold words are perceptually salient (i.e. stand out), but are not particularly legible as a font. Switching from roman to italic, however, does not slow down word recognition and suggests that words set in italic will therefore not function as well as bold for emphasis. Bold seems to be more appropriate than italic for setting headings or other access devices through making words stand out.

Typeface semantics

In Chapter 1 the idea of a typeface having a semantic role, as well as a functional role, was introduced. Typefaces can be suited to particular purposes not only because they are easy to read, but also because they convey a meaning though their visual form, sometimes described as personality. This is particularly relevant to marketing where brand names in appropriate typefaces (i.e. consistent with the product) are chosen more often than inappropriate ones (e.g. Doyle and Bottomley, 2004, 2006).

These two roles appear to be quite separate. A specific typeface might be more or less appropriate for a particular context (e.g. shop sign, wedding invitation, novel, textbook, annual report) but why would the legibility of this typeface be affected by its personality? This separation between legibility and aesthetics may not exist according to a captivating study. This study shows that we respond to words more slowly if the perceptual qualities of the font are inconsistent with the meaning of the word, e.g. the word ‘heavy’ in a ‘light’ font [8] (Lewis and Walker, 1989). Figure 5.23 illustrates words where the font is consistent or inconsistent with the meaning of the word. The origins of this effect are described in Box 5.9.

Consistent	**heavy**	*light*
Inconsistent	*heavy*	**light**

Figure 5.23 Two of the words used by Lewis and Walker (1989) set in Cooper Black (heavy) and Palatino Italic (light).

A more recent study confirms that using a font that is inconsistent with the word’s meaning (Figure 5.24) slows down decisions regarding the emotion conveyed (Hazlett, Larson, Shaikh, and Chaparro, 2013). Therefore, legibility can be influenced by the meaning conveyed by the typeface, although there does need to be quite a big difference between the personalities of the typefaces for this to emerge.

Consistent	*pretty*	reliable
Inconsistent	attractive	*proper*

Figure 5.24: Four of the words used by Hazlett, Larson, Shaikh, and Chaparro (2013) set in Corsiva and Times New Roman.

Box 5.9: Description and demonstration of the Stroop effect

Slowing responses when the font is inconsistent with the word's meaning is related to a well-known interference effect: the Stroop effect (Stroop, 1935/1992). The participant is asked to name the colour and is slower to respond if the word is inconsistent with the colour of the ink. Have a go at the demonstration in Figure 5.25.

blue	green	red	blue	brown	purple	brown	green	purple	red
purple	red	blue	brown	green	brown	purple	red	green	blue

Figure 5.25: Demonstration of the Stroop effect. Name the colour (not the word).

The tasks used in the two studies illustrated in Figures 5.22 and 5.23 are a reverse of the

Stroop effect as participants are asked to respond to the word's meaning (not the font).

Early studies described the connotations of typefaces as atmosphere value (Ovink, 1938) and congeniality (Zachrisson, 1970). The experimental approaches to determining the meaning have usually used semantic differential scales (see description of Rating in Panel 4.5). The dimensions that have emerged are:

- Evaluative measuring the value or worth of items (e.g. good versus bad; beautiful versus ugly)
- Potency measuring the strength (e.g. strong versus weak)
- Activity measuring action (e.g. active versus passive; fast versus slow)
- Mood measuring happiness (e.g. happy versus sad; relaxed versus tense)

The first three dimensions apply to many different types of things, e.g. political parties or works of art, but mood has been found to be particularly relevant to typefaces.

As typeface preferences and use change over time, it may be more helpful to look at results from studies in terms of more general patterns rather than the personalities of specific typefaces. Shaikh and Chaparro (2016) report an online survey of 40 on-screen typefaces with trends showing:

- Display typefaces that are bold, dark, block-like are viewed as stronger, less valuable, and more active (Broadway, Agency, Playbill)
- Script typefaces are seen as less strong, more valuable, and less active (Vivaldi, French Script, Monotype Corsiva)

But we should note that there are also individual typefaces within a category that deviate from these trends. Figure 5.26 illustrates the typefaces.

Display	Broadway	Agency	Playbill
Script	Vivaldi	French Script	Corsiva

Figure 5.26: Display typefaces viewed as stronger, less valuable, and more active. Script typefaces viewed as less strong, more valuable, and less active (Shaikh and Chaparro, 2016).

As we normally focus on reading, rather than examining the typeface, we may not be conscious of typeface connotations. But if asked to judge the appropriateness of a typeface for a particular type of text (e.g. professional or friendly), readers are aware of consistencies or inconsistencies (Brumberger, 2003).

We might expect typographers and graphic designers to be rather more focused on the personality of typefaces. A couple of studies have found some differences as to how the semantic qualities of typefaces are perceived based on the level of experience of design, but non-designers are able to perceive typeface connotations (Tannenbaum, 1964). There is quite a lot of agreement between designers and non-designers but there can also be pronounced differences on specific typefaces (Bartram, 1982). For example, designers rate Futura as positive on the Evaluative and Mood dimensions (e.g. beautiful, pleasant, good, happy, relaxed) whereas non-designers rate Futura as negative on these same dimensions (e.g. ugly, unpleasant, bad, sad, tense). Some caution should therefore be taken in assuming that your own perceptions will be a perfect match with all readers' perceptions.

?

Question: How would you go about checking that your choice of typeface(s) for a project is perceived as appropriate by the readers?

Rather than determining the meaning of a typeface directly, a few studies have looked at how the content of a text may be influenced by the typeface. Satirical articles on government issues and education policy set in Times New Roman were perceived as more satirical (angry and funny) than the same texts in Arial (Juni and Gross, 2008). However, this was not a very strong effect and an earlier study failed to show that the typeface can influence how the text content was perceived (Brumberger, 2003).

In the context of a job application, consideration should be given to the choice of typeface. Three identical resumés (CVs) set in three different typefaces (see Figure 5.27) can affect how an applicant is perceived (Shaikh and Fox, 2008).

Figure 5.27: The three typefaces used for CVs (Shaikh and Fox, 2008)

Question: Would you use any of these typefaces for your CV? If not, why not? Which of these typefaces would lead you to judge an applicant as knowledgeable, mature, experienced, professional, believable, and trustworthy? [9]

Despite the relevance of typeface connotations to choosing a typeface for a specific purpose, legibility is more important as a criterion of appropriateness than consistency for text-heavy document types (Shaikh and Chaparro, 2016). Readers are aware of the value of ease of reading.

Letter spacing

In Chapter 2 I introduced ‘crowding’ which refers to the effect of surrounding letters in words on the ease of identifying letters. It is easier to identify single letters if they are not in a word because the adjacent letters can jumble the appearance of letters. This suggests that increasing the space between letters would improve word identification, due to making it easier to identify the individual letters and accurately locate their position in relation to one another (e.g. to avoid confusing casual and causal). But if too much space is used, the letters may not be perceived as a group, i.e. a word. Another possible disadvantage of increasing letter spacing is that upcoming words, those we are about to read that are in our parafoveal vision, will be even further away from our fixation and therefore our acuity will be reduced.

A font has character widths (including space) built into it and most text processing software will have some way of adjusting the letter spacing from the default (0) value. Research exploring deviations from the defaults has produced the consistent finding that tightening (decreasing) letter spacing makes reading more difficult. However, increasing spacing has resulted in contradictory results: either no benefits or some benefit. This divergence can be explained by differences in readers, typefaces, method of testing, and amount of spacing.

Studies which use Courier, a monospaced typeface, found that tighter than standard spacing reduced reading speed but increasing beyond the standard did not increase reading speed (Chung, 2002; Yu, Cheung, Legge, and Chung, 2007). There was no evidence that the effect of crowding was reduced by increasing beyond the standard spacing. Using Courier is an odd choice from a designer’s perspective but it is easier to specify and manipulate space with each letter occupying the same fixed width.

However, this property may be the reason why there is no advantage to increasing letter spacing beyond standard. Monospacing (fixed width) results in looser spacing, particularly for narrower letters, and therefore might not need additional spacing, particularly as this means that words will extend further into our peripheral vision (Figure 5.28). The adjustments to letter spacing are also rather large compared with later studies.

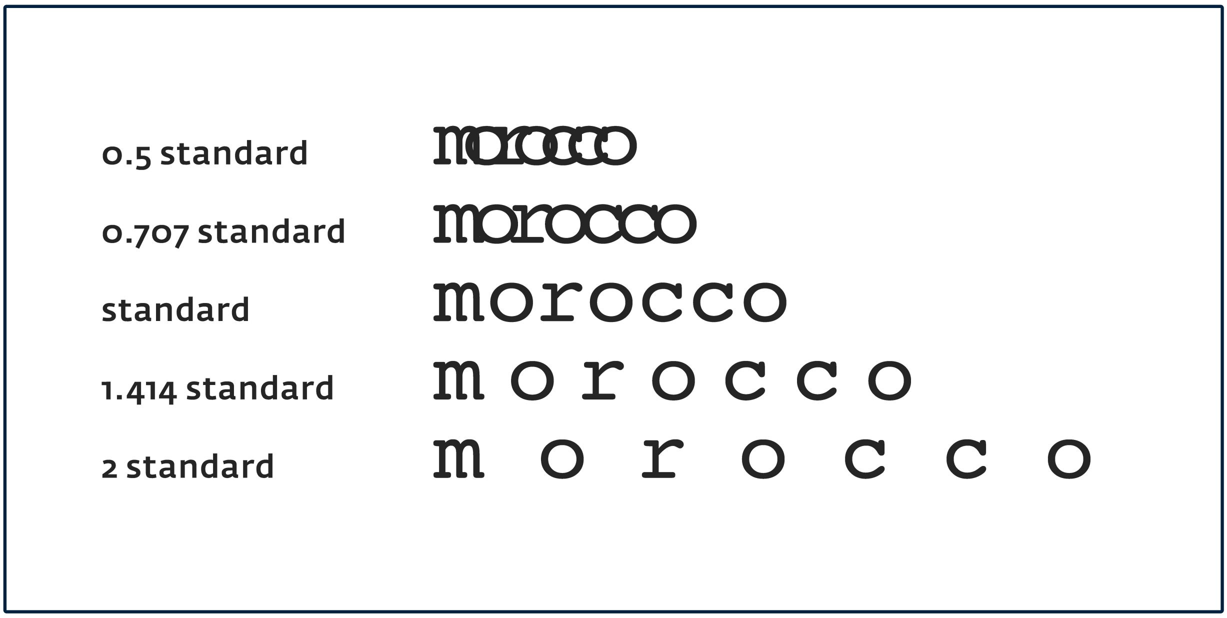


Figure 5.28: The monospaced typeface Courier used by Chung (2002). The standard spacing appears quite loose.

More recent studies have included adult skilled readers, young readers (7–8 and 9–10 year olds) and young readers with developmental dyslexia. Words set in 14 point Times New Roman with additional spacing (see Figure 5.29) are identified faster than the default spacing. This is true for all three sets of readers. However, when a reading task is used (not just single words) the advantage for wider spacing is only found with dyslexic readers. The reason why these different groups of readers were compared is because crowding tends to be greater for younger readers compared to adults and greater for dyslexics than normal readers (Perea, Moret-Tatay, and Gómez, 2011; Perea, Panadero, Moret-Tatay, and Gómez, 2012).

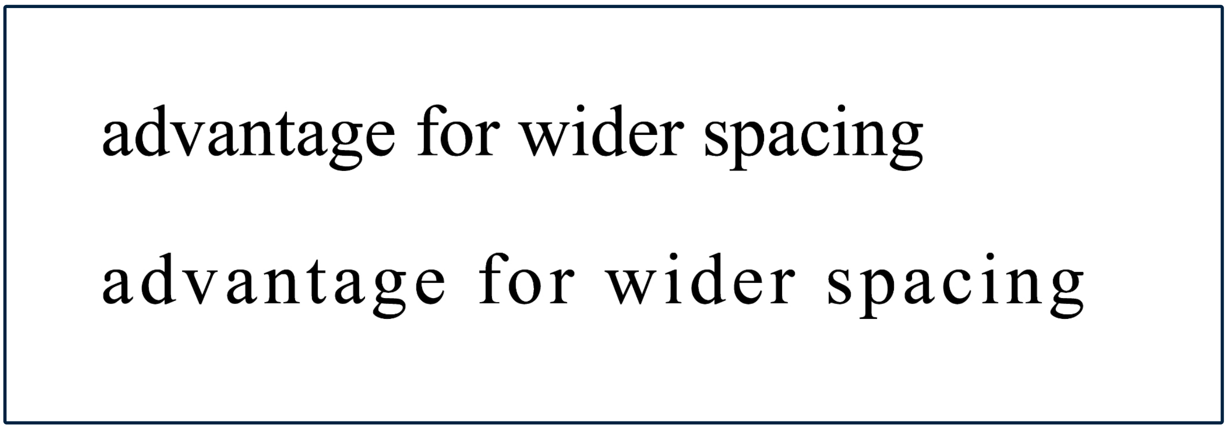


Figure 5.29 Default spacing (top) and wider spacing (bottom), described as increasing the value in Microsoft Word to 1.2 (Perea et al., 2011, 2012).

This relationship between increased crowding and spacing has been further explored by measuring the extent of crowding in individuals and looking at the corresponding effect of increasing letter, word, and line spacing (Joo, White, Strodtman, and Yeatman, 2018). Normal spacing consisted of words set in 11 point Calibri and spacing was increased by using 11 point Fluent Calibri (see Figure 5.30). A sub-group of adults with dyslexia who showed greater effects of crowding read faster with the additional spacing. This study did not determine whether letter, word, or line spacing is responsible for improving reading performance.

When 5–7 year old children were tested in a more natural reading environment, changes in letter spacing from tight to very wide had no effect on reading rate or errors (Reynolds and Walker, 2004). The children were asked to read aloud from a text set in 19 point Century Educational typeface which is used by publishers of early readers children's books.

With adult readers and more subtle changes in letter spacing (see Figure 5.31), responses get faster as spacing increases. Reading sentences, we fixate for a shorter time if there is more spacing (Perea and Gómez, 2012a, 2012b). However, when adults read in a more natural context (reading stories for comprehension), there is no difference in overall reading time between default spacing and expanded spacing (1.2), as in Figure 5.29 (Perea, Giner, Marcet, and Gomez, 2016). Although fixations are shorter with the extra space between letters, slightly more fixations are made which cancels out the advantage. The saccade length is similar in the default and expanded spacings. As we are used to reading the default setting, and initiating saccades of a specific length, the question remains as to whether we might adjust the number of fixations if we read the expanded text for longer.

Research has confirmed that relatively small adjustments to letter spacing will affect single word recognition in different ways depending on whether the typeface is proportional or fixed width (Slattery, Yates, and Angele, 2016). When letter spacing is increased:

- Words in proportional width typefaces (Calibri, Cambria, Georgia, and Verdana) are responded to faster.
- Words in fixed width typefaces (Consolas and Courier New) are responded to slower. This confirms the earlier studies described above which also used Courier.

This study found no differences between serif and sans serif typefaces. This may be because the default spacings are adjusted appropriately (see Figure 5.32).

Figure 5.30 Examples of word lists used by (Joo et al., 2018) with normal spacing in Calibri (top) and increased spacing in Fluent Calibri (bottom).

more subtle changes in letter spacing

more subtle changes in letter spacing

more subtle changes in letter spacing

more subtle changes in letter spacing

more subtle changes in letter spacing

Figure 5.31: Interletter spacing of -0.5 (condensed), 0 (default), +0.5, +1.0, +1.5 (expanded) used by Perea and Gómez (2012a).

Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?

Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?

Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?

Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?

Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?

Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?
Does letter spacing depend on the typeface?

Figure 5.32: Comparison of proportional and fixed width typefaces with three letter spacings: tighter than default, default, and looser than default. This range of letter spacing is similar to that used by Slattery, Yates and Angele ([2016](/en/bibliography#yates-angele-2016)). The typefaces are (top to bottom) Calibri, Cambria, Georgia, Verdana, Consolas and Courier New.

In summary, there is some evidence that slightly looser letter spacing helps dyslexic readers and possibly other readers, but this will depend on the typeface. E-books might therefore benefit from including control over letter spacing, particularly for dyslexic readers. Adults spend less time on individual fixations with looser spacing but more fixations are made. It therefore looks as though the defaults determined by type designers are appropriate even though they are not based on empirical research.

Word spacing

In the studies above where sentences were read, word spacing increased when letter spacing increased. Microsoft Word also automatically adjusts spaces between words when the user changes letter spacing (see, for example, Figure 5.31).

Word spaces allow us to segment text into words and help us target where to land our eyes, based on parafoveal vision. The space before the first letter of a word and space after the last letter also reduce crowding effects. All of these factors argue for having quite large word spaces. But there is a disadvantage as increasing word spaces pushes words further into our peripheral vision where our acuity drops off sharply. If letter spacing is reduced but word spacing increased, upcoming words are not pushed further into peripheral vision (see Figure 5.33). One study which used these spacings with Georgia and Consolas (a fixed width sans serif) found a benefit from reducing letter spacing (a little) and increasing word spacing. However, this was mainly with Georgia, rather than Consolas, which can be explained by the default spacings (see Figure 5.32). Georgia has tighter word spacing and so can benefit more from an increase than Consolas (Slattery and Rayner, 2013).

Relationship between letter and word spacing can be adjusted

Relationship between letter and word spacing can be adjusted

`Relationship between letter and word spacing can be adjusted`

`Relationship  between  letter  and  word  spacing  can  be  adjusted`

Figure 5.33: The top sentence in each pair has the default letter and word spacing. The bottom sentence in each pair has reduced letter spacing and increased word spacing. The top pair uses the typeface Georgia and the bottom pair is in Consolas.

This relationship between word and letter spacing was explored further by Slattery, Yates, and Angele (2016) using Calibri and Consolas to compare proportional and fixed width typefaces. They confirm the importance of considering letter and word spacing together and propose that spaces between words should be at least 3.5 times the spaces between letters for efficient reading. As with letter spacing, fixed width Consolas is read slower if word spacing is larger than default, unless letter spacing is reduced. Calibri is read slower if word spacing is reduced. Although this study involved reading sentences, the researchers point out that only single lines of text were used and more research is required using multiple lines.

In summary, when making adjustments to word spacing:

- defaults built into different fonts and page layout software need to be taken into account
- the relationship between letter and word spacing is important
- characteristics of readers may be particularly relevant as younger readers and dyslexics are more susceptible to crowding effects

Alignment

Paragraphs of text are typically aligned on the left with the right margin either aligned to produce justified text or with a ragged right margin to produce unjustified text. Most studies have shown no differences in reading speed although fully justified text may be problematic for poor readers when set in short lines, i.e. 7 words per line, about 42 characters (see Figure 5.34) or even slightly longer, about 52 characters (Gregory and Poulton, 1970; Zachrisson, 1965). One study has looked at alignment in web pages but using a search for a link in a screen of text, rather than reading the text. Performance was better with left-aligned then justified text although participants preferred justified (Ling and van Schaik, 2007).

Paragraphs of text are typically aligned on the left with the right margin either aligned to produce justified text or uneven to produce unjustified text. The length of line will influence the extent to which justification affects word spacing and/or hyphenation. Shorter lines will result in more variable word spacing especially when words are long and no hyphenation is used.

Paragraphs of text are typically aligned on
the left with the right margin either
aligned to produce justified text or uneven
to produce unjustified text. The length of
line will influence the extent to which
justification affects word spacing and/or
hyphenation. Shorter lines will result in
more variable word spacing especially
when words are long and no hyphenation
is used.

Figure 5.34: Short lines set in justified (top) and unjustified (bottom) setting.

The main reason proposed for the reduced legibility of justified text is the uneven spacing, often described as 'rivers' of white. These larger word spaces are more likely to occur with:

- short lines
- lots of long words
- wider characters
- no hyphenation
- less sophisticated control over word (and letter) spacing

The assumption is that eye movements will be adversely affected by this unevenness which might be due to the lack of rhythm. Another possible explanation is that larger word spaces push forthcoming words further into peripheral vision, reducing their acuity. These factors may be more important for poorer readers who have a smaller perceptual span (and who therefore make use of fewer letters to the right). These proposed explanations have not been tested, as far as I know. It is possible that the issue is not differences in legibility but aesthetic considerations.[10]

Line length

Line length, sometimes described as line width, can be measured by:

- physical length of the line (e.g. 15 centimetres), sometimes converted to visual angle by taking into account the viewing distance
- number of characters per line (cpl) which can be varied by
  1) changing type size, keeping physical length constant, or
  2) keeping type size constant which varies physical length

  Figure 5.35 shows the various ways of changing line lengths.

Line length can be measured by physical length of the line (e.g. 15 centimetres), converted to visual angle by taking into account the viewing distance or number of characters per line which can be varied by changing type size, keeping physical length constant, or keeping type size constant which varies physical length.

Line length can be measured by physical length of the line (e.g. 15 centimetres), converted to visual angle by taking into account the viewing distance or number of characters per line which can be varied by changing type size, keeping physical length constant, or keeping type size constant which varies physical length.

Line length can be measured by physical length of the line (e.g. 15 centimetres), converted to visual angle by taking into account the viewing distance or number of characters per line which can be varied by changing type size, keeping physical length constant, or keeping type size constant which varies physical length.

Figure 5.35: Examples showing the relationship between physical line length, number of characters per line, and type size. Top and middle: same physical length but smaller type size in middle increases number of characters per line. Top and bottom: same number of characters per line but physical length varies because of type size.

Research into the relative legibility of different line lengths in print has led to recommendations that line lengths should not exceed about 70 cpl (Spencer, 1968). Various studies summarised in Tinker (1963) have been interpreted as supporting an optimal line length of 52 cpl (Rayner and Pollatsek, 1989, p118). The explanation given for the legibility of this moderate line length is that it is the outcome of a trade-off between two opposing factors. If line lengths are too long, the return sweeps to the beginning of the next line can be inaccurate. If the lines are too short, readers cannot make maximum use of their peripheral vision. However, recent studies have questioned whether inaccurate return sweeps are necessarily problematic.

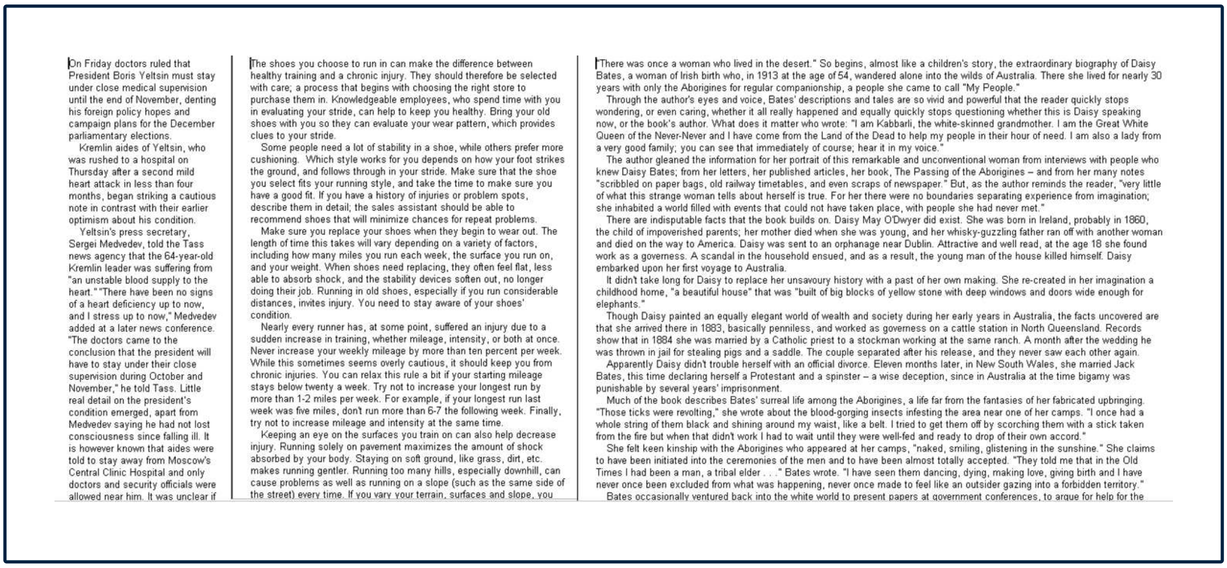
Studies looking at line length on screen began in the 1980s but were constrained as to how they could change line length. They did this by changing the character density which meant characters of the same height but different widths, looking something like Figure 5.36.

Character density can be considered as a change in type size except characters change width and not height. This is simulated here by condensing characters.

Character density can be considered as a change in type size except characters change width and not height. This is simulated here by condensing characters.

Figure 5.36: A simulation of different character densities used in the 1980s to change type size and therefore number of characters per line.

An early study showed that smaller characters, with more characters per line (bottom of Figure 5.36) are read faster and more efficiently with fewer fixations overall (Kolers, Duchnicky, and Ferguson, 1981). The line lengths compared were 35 and 70 cpl. Around 20 years later, using updated screen technology, line length was varied by changing the number of characters of the same size (Figure 5.37). The study found that reading rate tended to get faster as lines got longer, surprisingly even up to 100 cpl (Dyson, 2004, 2005). Similar advantages for longer lines on screen were found when searching for words, rather than continuous reading (Youngman and Scharff, 1998; Ling and van Schaik, 2006).



On Friday doctors ruled that President Boris Yeltsin must stay under close medical supervision until the end of November, denting his foreign policy hopes and campaign plans for the December parliamentary elections.

Kremlin aides of Yeltsin, who was rushed to a hospital on Thursday after a second mild heart attack in less than four months, began striking a cautious note in contrast with their earlier optimism about his condition.

Yeltsin's press secretary, Sergei Medvedev, told the Tass news agency that the 64-year-old Kremlin leader was suffering from "an unstable blood supply to the heart." "There have been no signs of a heart deficiency up to now, and I stress up to now," Medvedev added at a later news conference. "The doctors came to the conclusion that the president will have to stay under their close supervision during October and November," he told Tass. Little real detail on the president's condition emerged, apart from Medvedev saying he had not lost consciousness since falling ill. It is however known that aides were told to stay away from Moscow's Central Clinic Hospital and only doctors and security officials were allowed near him. It was unclear if

The shoes you choose to run in can make the difference between healthy training and a chronic injury. They should therefore be selected with care; a process that begins with choosing the right store to purchase them in. Knowledgeable employees, who spend time with you in evaluating your stride, can help to keep you healthy. Bring your old shoes with you so they can evaluate your wear pattern, which provides clues to your stride.

Some people need a lot of stability in a shoe, while others prefer more cushioning. Which style works for you depends on how your foot strikes the ground, and follows through in your stride. Make sure that the shoe you select fits your running style, and take the time to make sure you have a good fit. If you have a history of injuries or problem spots, describe them in detail; the sales assistant should be able to recommend shoes that will minimize chances for repeat problems.

Make sure you replace your shoes when they begin to wear out. The length of time this takes will vary depending on a variety of factors, including how many miles you run each week, the surface you run on, and your weight. When shoes need replacing, they often feel flat, less able to absorb shock, and the stability devices soften out, no longer doing their job. Running in old shoes, especially if you run considerable distances, invites injury. You need to stay aware of your shoes' condition.

Nearly every runner has, at some point, suffered an injury due to a sudden increase in training, whether mileage, intensity, or both at once. Never increase your weekly mileage by more than ten percent per week. While this sometimes seems overly cautious, it should keep you from chronic injuries. You can relax this rule a bit if your starting mileage stays below twenty a week. Try not to increase your longest run by more than 1-2 miles per week. For example, if your longest run last week was five miles, don't run more than 6-7 the following week. Finally, try not to increase mileage and intensity at the same time.

Keeping an eye on the surfaces you train on can also help decrease injury. Running solely on pavement maximizes the amount of shock absorbed by your body. Staying on soft ground, like grass, dirt, etc. makes running gentler. Running too many hills, especially downhill, can cause problems as well as running on a slope (such as the same side of the street) every time. If you vary your terrain, surfaces and slope, you

"There was once a woman who lived in the desert." So begins, almost like a children's story, the extraordinary biography of Daisy Bates, a woman of Irish birth who, in 1913 at the age of 54, wandered alone into the wilds of Australia. There she lived for nearly 30 years with only the Aborigines for regular companionship, a people she came to call "My People."

Through the author's eyes and voice, Bates' descriptions and tales are so vivid and powerful that the reader quickly stops wondering, or even caring, whether it all really happened and equally quickly stops questioning whether this is Daisy speaking now, or the book's author. What does it matter who wrote: "I am Kabbarli, the white-skinned grandmother. I am the Great White Queen of the Never-Never and I have come from the Land of the Dead to help my people in their hour of need. I am also a lady from a very good family; you can see that immediately of course; hear it in my voice."

The author gleaned the information for her portrait of this remarkable and unconventional woman from interviews with people who knew Daisy Bates; from her letters, her published articles, her book, The Passing of the Aborigines – and from her many notes "scribbled on paper bags, old railway timetables, and even scraps of newspaper." But, as the author reminds the reader, "very little of what this strange woman tells about herself is true. For her there were no boundaries separating experience from imagination; she inhabited a world filled with events that could not have taken place, with people she had never met."

There are indisputable facts that the book builds on. Daisy May O'Dwyer did exist. She was born in Ireland, probably in 1860, the child of impoverished parents; her mother died when she was young, and her whisky-guzzling father ran off with another woman and died on the way to America. Daisy was sent to an orphanage near Dublin. Attractive and well read, at the age 18 she found work as a governess. A scandal in the household ensued, and as a result, the young man of the house killed himself. Daisy embarked upon her first voyage to Australia.

It didn't take long for Daisy to replace her unsavoury history with a past of her own making. She re-created in her imagination a childhood home, "a beautiful house" that was "built of big blocks of yellow stone with deep windows and doors wide enough for elephants."

Though Daisy painted an equally elegant world of wealth and society during her early years in Australia, the facts uncovered are that she arrived there in 1883, basically penniless, and worked as governess on a cattle station in North Queensland. Records show that in 1884 she was married by a Catholic priest to a stockman working at the same ranch. A month after the wedding he was thrown in jail for stealing pigs and a saddle. The couple separated after his release, and they never saw each other again.

Apparently Daisy didn't trouble herself with an official divorce. Eleven months later, in New South Wales, she married Jack Bates, this time declaring herself a Protestant and a spinster – a wise deception, since in Australia at the time bigamy was punishable by several years' imprisonment.

Much of the book describes Bates' surreal life among the Aborigines, a life far from the fantasies of her fabricated upbringing. "Those ticks were revolting," she wrote about the blood-gorging insects infesting the area near one of her camps. "I once had a whole string of them black and shining around my waist, like a belt. I tried to get them off by scorching them with a stick taken from the fire but when that didn't work I had to wait until they were well-fed and ready to drop of their own accord."

She felt keen kinship with the Aborigines who appeared at her camps, "naked, smiling, glistening in the sunshine." She claims to have been initiated into the ceremonies of the men and to have been almost totally accepted. "They told me that in the Old Times I had been a man, a tribal elder . . ." Bates wrote. "I have seen them dancing, dying, making love, giving birth and I have never once been excluded from what was happening, never once made to feel like an outsider gazing into a forbidden territory."

Bates occasionally ventured back into the white world to present papers at government conferences, to argue for help for the

Figure 5.37: Three line lengths used by Dyson in various studies which change the number of characters, keeping type size (and line spacing) constant. The top is 25 cpl, the middle 55 cpl, and the bottom 100 cpl.

However, a consistent finding is that long line lengths on screen are least preferred or judged as least easy to read. Moderate line lengths (around 50–70 cpl) are preferred which fits with preferences for line lengths in print. This makes sense as readers may make similar judgements regardless of whether they are reading print or from screen. Readers may also be judging what they frequently encounter in printed material as being easiest to read, i.e. familiarity determining perceived legibility. But why might there be an advantage in reading speed for long lines (up to around 100 characters) on screen whereas in print, a maximum of 70 characters is most legible? Here are some suggestions:

- The subjective judgements might provide a hint. Long line lengths on screen look quite daunting and this may encourage faster less detailed reading, such as scanning or skimming. This is consistent with comprehension being poorer with long line lengths, compared with moderate length lines. In fact, Ling and van Schaik (2006) suggest that longer line lengths should be used for quick scanning and shorter lines when text needs to be read more thoroughly.
- The mechanics of reading texts on screen might also influence reading time. If

required to scroll through texts, a longer line length will require less scrolling as there will be fewer lines. If we read whilst scrolling, we won't be slowed down by greater amounts of scrolling. If we don't, long line lengths will be more efficient to navigate.[11]

- A question remains as to why long lines might be less problematic in terms of the accuracy of the return sweep from the end of one line to the beginning of the next. As mentioned above, we tend to sit further away from a desktop computer screen than printed material which means that a line of text has a smaller visual angle. This smaller visual angle may make it easier to locate the correct position for our eyes when we do a return sweep. Another possible explanation is that the scrolling provides a cue to locating the next line, the upward movement of text reducing the difficulty with long lines. However, this is only likely to work if we scroll slowly, and may therefore be less relevant to touchscreen interfaces used on smartphones or tablets.

Question: Are you convinced by any of the above suggestions? If you encounter a long line length on the screen of a desktop computer, do you adjust the settings? If so, why do you do this? Are you influenced by what you read in print?

Columns

Another means of varying line length is to set text in columns: multiple columns generally result in shorter line lengths. There are a few studies which have directly compared single and multiple columns in print. The context for much of this research was exploring academic printed journal designs which typically use multiple column formats. The findings were not entirely consistent:

- an advantage for narrow column setting (Foster, 1970)
- single columns read faster than double columns (Poulton, 1959)
- no difference between a single column and double column (Hartley, Burnhill, and Fraser, 1974)
- with children aged around 11–12 years old there was slightly faster scanning for items in two columns (around 53 cpl) compared to one column (around 115 cpl). However, the longer line length was not problematic (Hartley, Burnhill, and Davis, 1978).

In the early days of online versions of newspapers and magazines, column formats were typically used on the web, often as PDFs, until these were re-designed and tailored for screen viewing. In a comparison of one and three columns on screen (Figure 5.38), the single column (80 cpl) was read faster with no differences in comprehension (Dyson and Kipping, 1997#dyson-kipping-1997). This provides further support for the advantage of longer lines described above.[12]

Some interesting additional findings from the Dyson and Kipping study are:

- The advantage for this single column was restricted to a younger age group (18–24 year olds). Those over 25 years old read each version at a similar speed. We surmised that the younger participants were more familiar with reading web pages (around 1997).
- Comprehension was better for faster readers in the three-column version than for slower readers. We speculated that faster readers may use a different reading pattern and be able to scan narrow columns in an efficient manner to absorb information.

When I was seven I found a machine gun under the Christmas tree. In my small hands it so excited me that I immediately forgot my unopened presents and began shooting my brother and sister. The noise and smoke threw the room into a state of confusion. My brother uncovered a pair of six-shooters and returned fire. By the time my parents got us under control an acrid haze filled the living room and my sister was bawling about how Christmas had been ruined for everyone.

Over the years there have been many grass-roots attempts to take violent toys out of the hands of children, largely without success. Idealistic parents have discovered that if you take a little boy's gun away, he will probably pick up a stick and continue the game. Most have reluctantly accepted war games as a normal part of growing up — or simply ignored the issue.

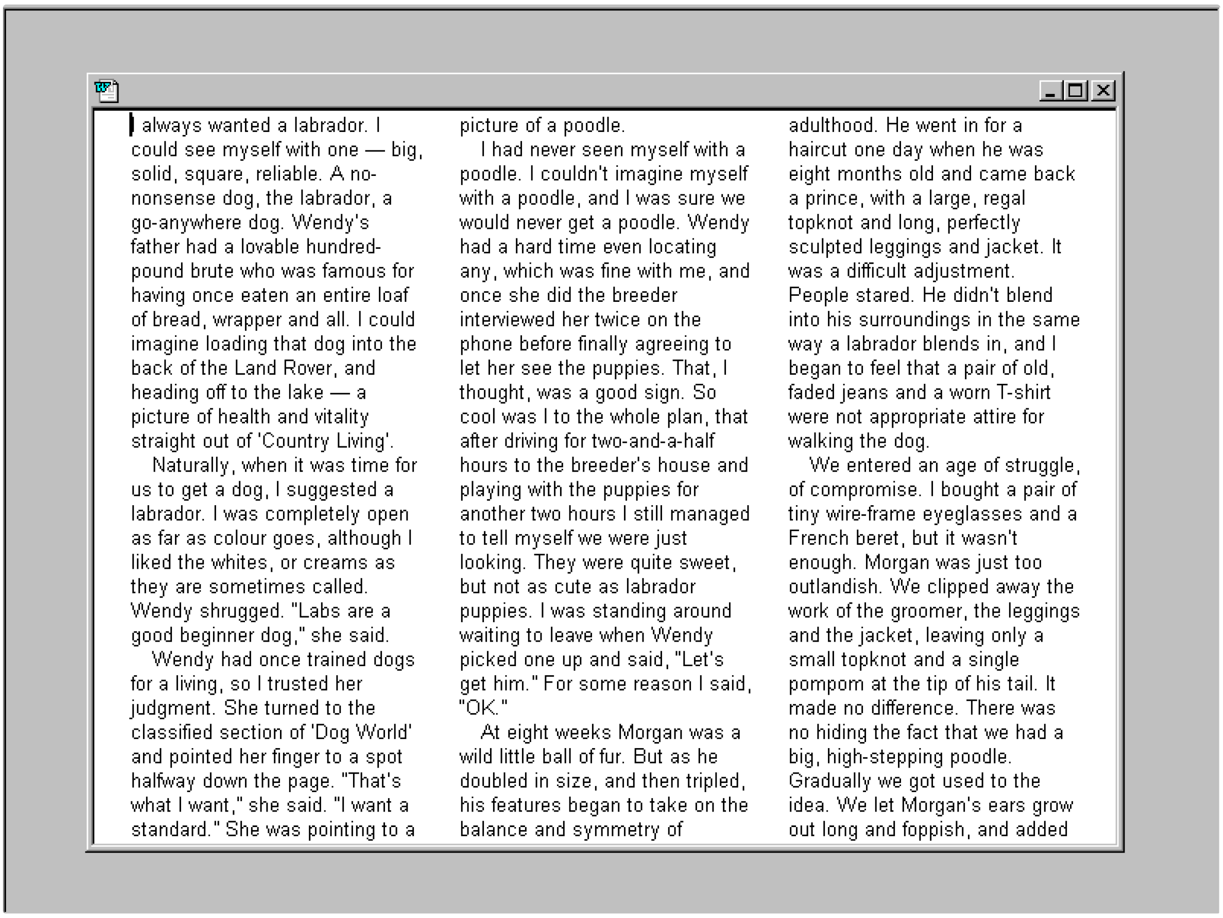
But the graphic mayhem of modern electronic games has reopened the debate. It is indeed shocking to witness, in games like Doom and Mortal Kombat, such carnage as we with our sticks and plastic guns could only imagine. And this is only the beginning. As the technology improves, the enemies your child eviscerates will become more and more lifelike.

This is natural in a free-market economy. Video-game manufacturers operate under the cold-war dictum that if you don't produce it, your competitor will. And so, like fearful superpowers, they continue to amass their electronic instruments of destruction.

Of course we don't live in a totally free market. Certain things are forbidden. You can no longer buy cocaine or heroin at the corner drug store, you cannot buy hand grenades or surface-to-air missiles, you can't buy a packet of plutonium for your children's science project. These prohibitions are for our own good. Alcohol was once prohibited for our own good also, but sometimes the government goes too far. Alcohol is now labeled, as are cigarettes: danger, if this kills you, don't blame us.

We now have labels on video and computer games. Unlike cigarette labels, video-game rating labels are 'voluntary.' There are actually two rating systems, one adopted by the 'Interactive Digital Software Association' and one adopted by the 'Software Publishers of America'. Since they are voluntary labels, they are naturally discreet and almost meaningless. You pick up an IDSA game and see a small sticker that says, 'Teens.' What does that mean? A lot of violence? A little violence? If I were a teen I'd probably go for the box labeled 'Mature,' just to be safe. An SPA game will have a sticker with the word 'Violence,' and next to it a small thermometer like you see on a can of chili peppers. I found myself drawn to the hot ones. Hot equals exciting.

These stickers will probably change nothing. The pressure within the industry to produce increasingly



I always wanted a labrador. I could see myself with one — big, solid, square, reliable. A no-nonsense dog, the labrador, a go-anywhere dog. Wendy's father had a lovable hundred-pound brute who was famous for having once eaten an entire loaf of bread, wrapper and all. I could imagine loading that dog into the back of the Land Rover, and heading off to the lake — a picture of health and vitality straight out of 'Country Living'.

Naturally, when it was time for us to get a dog, I suggested a labrador. I was completely open as far as colour goes, although I liked the whites, or creams as they are sometimes called. Wendy shrugged. "Labs are a good beginner dog," she said.

Wendy had once trained dogs for a living, so I trusted her judgment. She turned to the classified section of 'Dog World' and pointed her finger to a spot halfway down the page. "That's what I want," she said. "I want a standard." She was pointing to a picture of a poodle.

I had never seen myself with a poodle. I couldn't imagine myself with a poodle, and I was sure we would never get a poodle. Wendy had a hard time even locating any, which was fine with me, and once she did the breeder interviewed her twice on the phone before finally agreeing to let her see the puppies. That, I thought, was a good sign. So cool was I to the whole plan, that after driving for two-and-a-half hours to the breeder's house and playing with the puppies for another two hours I still managed to tell myself we were just looking. They were quite sweet, but not as cute as labrador puppies. I was standing around waiting to leave when Wendy picked one up and said, "Let's get him." For some reason I said, "OK."

At eight weeks Morgan was a wild little ball of fur. But as he doubled in size, and then tripled, his features began to take on the balance and symmetry of adulthood. He went in for a haircut one day when he was eight months old and came back a prince, with a large, regal topknot and long, perfectly sculpted leggings and jacket. It was a difficult adjustment. People stared. He didn't blend into his surroundings in the same way a labrador blends in, and I began to feel that a pair of old, faded jeans and a worn T-shirt were not appropriate attire for walking the dog.

We entered an age of struggle, of compromise. I bought a pair of tiny wire-frame eyeglasses and a French beret, but it wasn't enough. Morgan was just too outlandish. We clipped away the work of the groomer, the leggings and the jacket, leaving only a small topknot and a single pompom at the tip of his tail. It made no difference. There was no hiding the fact that we had a big, high-stepping poodle. Gradually we got used to the idea. We let Morgan's ears grow out long and foppish, and added

Figure 5.38: Comparison of a single column (80 cpl) and three columns (each of 25 cpl) used by Dyson and Kipping (1997).

Line spacing

A very early study indicated that increasing line space from no additional space (same as point size) to 7 points additional space led to faster reading; with more than 7 points, reading slowed down (Bentley, 1921). As the type sizes used were 6, 9 and 12 point, this is rather generous use of line space (see Figure 5.39).

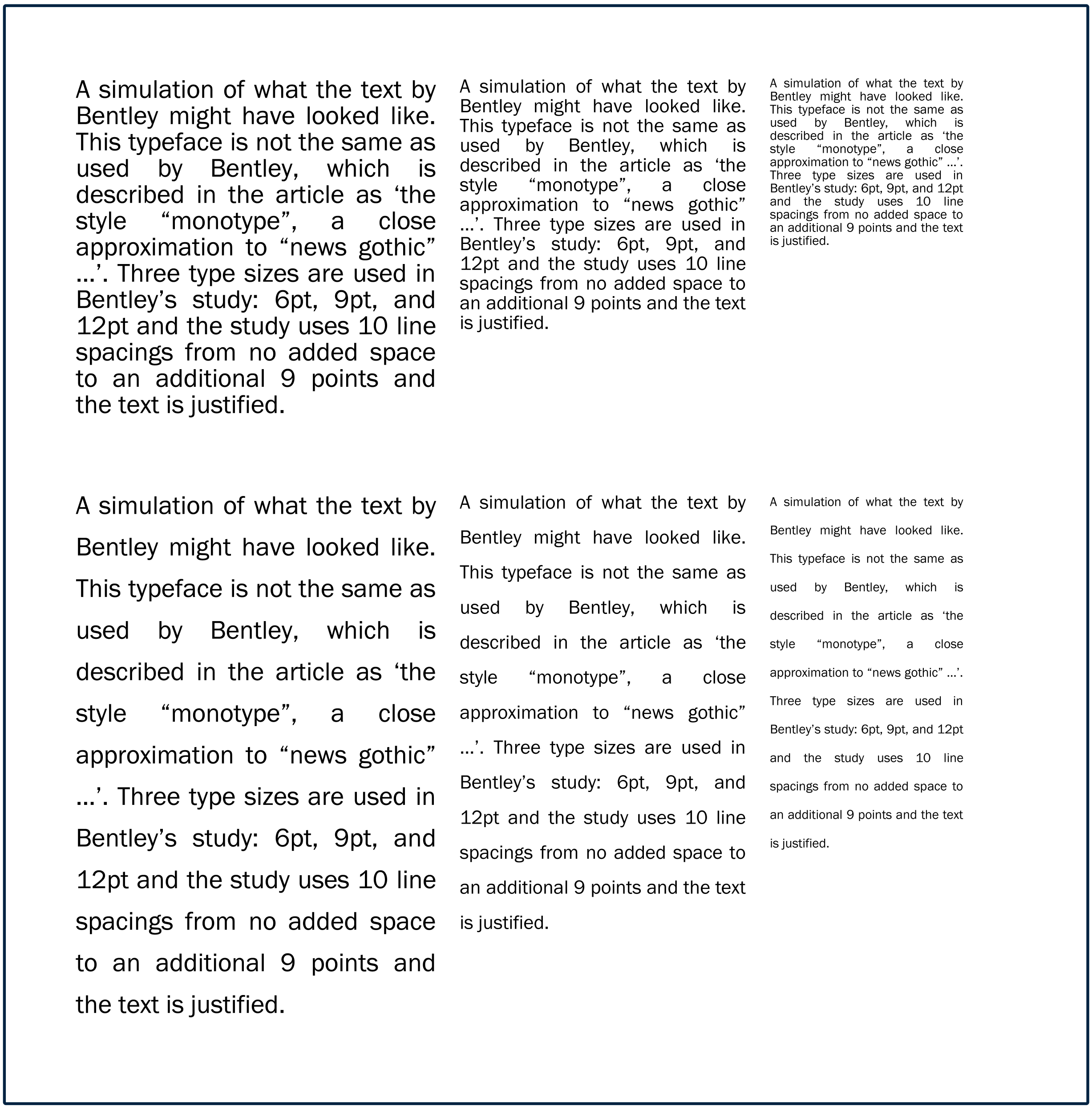


Figure 5.39: An indication of the material used by Bentley (1921). The top row shows the three type sizes (12, 9 and 6 point) with no additional line space. The bottom row has the same sizes with 7 points additional space.

As discussed in Chapter 4, designers do not make decisions on individual typographic variables, but consider the relationship between these variables. The legibility of a particular line length may be influenced by the amount of line spacing. The problem of inaccurate return sweeps when reading a long line may be alleviated by introducing more space between the lines. Paterson and Tinker studied type size, line length and line spacing of print by systematically varying all three (summarised in Tinker, 1965). The results were expressed as ‘safety zones’ referring to limits of line length and line spacing within which legibility would be satisfactory. These were generally between 1 and 4 points with not too much variation according to line length or type size. What we may conclude from this is that line spacing should not be too tight. Different typefaces are also perceived as needing different amounts of line spacing to maximise their attractiveness or appeal for reading with sans serif and italic needing one point more than serif roman typefaces (Becker, Heinrich, von Sichowsky, and Wendt, 1970).

Such extensive research has not been carried out for reading from screen. Given the rather general outcomes, it is questionable as to whether it would be worth repeating for screen. One study with the objective of generating design guidelines for web pages compared Arial 10 point type set in single, 1.5 or double line spacing. The researchers found that the greater the line space, the better able participants were to locate hyperlinks within texts and their preferences also followed this pattern (Ling and van Schaik, 2007).

Locating hyperlinks is an information retrieval task which will not involve the same sequence of eye movements as continuous reading. It is plausible that words are more easily identified when there is more space above and below them. This seems to be confirmed by a study which looked at line spacing from the perspective of crowding (Chung, 2004). Crowding has been discussed above in terms of letter and word spacing, i.e. horizontal space, whereas this study looked at vertical space which designers call line spacing. The study looked at the speed of identifying words and the results indicated that an increase in space above and below words increased reading speed. If we look directly at the word so that it falls onto the fovea, we benefit from increases of about 1.25 to 1.5 the standard spacing (see Figure 5.40) and after that there is no additional benefit. However, for words in peripheral vision which we use to guide where we land our eyes following a saccade, even greater line spacing is beneficial. These effects are greater for smaller type sizes suggesting that line spacing should not be set at a consistent percentage of type size; small sizes need relatively more additional space. Although Tinker’s method and approach were very different, and the results expressed in a different way, there is some agreement with the crowding conclusions.

Figure 5.40: Variations in space above and below words (line space) used by Chung (2004). The study found that 1.25x to 1.5x the standard spacing (1x) increased the speed of identifying the middle word (bike).

These studies indicate that optimal line spacing may be similar in print and on screen. The benefits of generous spacing may be an aid to improving the accuracy of return sweeps in longer lines, but also a means of alleviating crowding from adjacent lines.

Paragraph denotation

The typical ways of denoting paragraphs are:

- space between paragraphs

- indented first line
- new line but no indent
- occasionally, indented first line and space between paragraphs

The first three of these formats plus no denotation have been compared with 11–12 year olds scanning a text for missing words (Hartley, Burnhill, and Davis, 1978). They found:

- the version with space between paragraphs (1 in Figure 5.41) is scanned faster than the new line with no indent (3) and no denotation (4)
- the version with additional space between paragraphs (1) is not scanned significantly faster than paragraphs denoted by an indent (2)

They conclude that as an indent uses less space, this is a more cost-effective solution for print.

There does not appear to be any research which has pursued ways of denoting paragraphs on screen. The cost is not a factor and typically space is used to separate paragraphs (as in 1).

1.
There are several ways to denote the beginning of a new paragraph using horizontal or vertical space. If a new line is the only cue to a new paragraph this can be problematic.

What happens in the last line of one paragraph fills the whole line? The next paragraph begins on the next line but this looks no different from continuing the same paragraph.

2.
There are several ways to denote the beginning of a new paragraph using horizontal or vertical space. If a new line is the only cue to a new paragraph this can be problematic.
    What happens in the last line of one paragraph fills the whole line? The next paragraph begins on the next line but this looks no different from continuing the same paragraph.

3.
There are several ways to denote the beginning of a new paragraph using horizontal or vertical space. If a new line is the only cue to a new paragraph this can be problematic.
What happens in the last line of one paragraph fills the whole line? The next paragraph begins on the next line but this looks no different from continuing the same paragraph.

4.
There are several ways to denote the beginning of a new paragraph using horizontal or vertical space. If a new line is the only cue to a new paragraph this can be problematic. What happens in the last line of one paragraph fills the whole line? The next paragraph begins on the next line but this looks no different from continuing the same paragraph.

Figure 5.41: Methods of denoting paragraphs used by Hartley, Burnhill, and Davis (1978): (1) additional space; (2) indented first line; (3) starting new line; (4) no denotation.

Question: How do you typically denote paragraphs in print and on screen? What criteria determine your decisions?

Headings

Headings have the function of structuring a text, signalling the topic covered in the following text, and providing an access point (i.e. to locate a particular section). To serve these purposes, headings need to be differentiated from surrounding text. Earlier in this chapter, bold was considered as more appropriate than italic as a means of making words stand out. Comparing all capitals and bold, newspaper headlines were found to be located faster in bold lower case than all capitals (Poulton, 1967). The lower case x-height was matched to the height of capital letters as the typographic designer involved in this study considered these to be optimal for the setting (see Figure 5.19). The capitals were set at the size that were in current use in the newspaper. This approach was in keeping with the practical purpose of the research which was carried out at the request of the editor of *The Times* newspaper.

Research comparing different graphic treatments of headings is however very limited and there has been more emphasis on the linguistic function of headings in facilitating processing of text and improving recall (e.g. Hyönä and Lorch, 2004). We do know that the position of headings (embedded or in the margin) did not matter for 14–15 year olds (Hartley and Trueman, 1983). (Figure 5.42) We might therefore conclude that the two solutions are similar in terms of usability.

Headings
Headings have the function of structuring a text, providing an access point (i.e. to locate a particular section) and signalling the topic covered in the following text. They help readers to search, recall and retrieve material from a text (Hartley and Trueman, 1983). To serve these purposes, headings need to be differentiated from surrounding text. Earlier in this chapter, bold was considered as more appropriate than italic as a means of making words stand out.

Research comparing different graphic treatments of headings is very limited and there has been more emphasis on the linguistic function of headings in facilitating processing of text and improving recall (e.g. Hyönä and Lorch, 2004). We do know that the position of headings (in margin or embedded) did not matter for 14-15 year olds (Hartley and Trueman, 1983).

Headings	Headings have the function of structuring a text, providing an access point (i.e. to locate a particular section) and signalling the topic covered in the following text. They help readers to search, recall and retrieve material from a text (Hartley and Trueman, 1983). To serve these purposes, headings need to be differentiated from surrounding text. Earlier in this chapter, bold was considered as more appropriate than italic as a means of making words stand out. Research comparing different graphic treatments of headings is very limited and there has been more emphasis on the linguistic function of headings in facilitating processing of text and improving recall (e.g. Hyönä and Lorch, 2004). We do know that the position of headings (in margin or embedded) did not matter for 14-15 year olds (Hartley and Trueman, 1983).

Figure 5.42: Two positions of headings used by Hartley and Trueman (1983): embedded (top) and in the margin (bottom).

One study took a different approach to identifying the most appropriate typography for headings in text by exploring how easy it is to visually discriminate among them using a set of cards (Williams and Spyridakis, 1992). They measured the time required to place 16 different heading treatments in order of importance. The assumption is that if we can do this quickly, this suggests that the headings are clearly different from each other and consequently, they would work in a text to indicate the hierarchical structure. The treatments used type size, position (centred, flush left, indented, embedded), underlining, and case. Type size was perceived as the most powerful cue to the hierarchical status of the heading, which is understandable as there is little ambiguity in this treatment: a larger heading means a higher level of heading. A more subtle finding is that their participants found it easier to make judgements when the headings varied along fewer dimensions (e.g. size alone versus size and position).

Some care should be taken in applying these results to practice as headings need to be differentiated from body text as well as from other headings.[13] It may therefore be desirable to change at least two dimensions to identify headings (e.g. size, case, typeface or type variant) and indicate the hierarchy of headings through one change, possibly size. There may be an interesting difference in which variable is used in different countries. In Mexico, upper case is frequently used for headings.

The space above and below headings appears not to have been specifically investigated in an empirical study but the Gestalt laws of grouping and organisation can inform practice (Panel 5.4). The Gestalt law of proximity states that elements positioned closer together are seen as a group and perceived to be more closely related than other elements in the image or display. Headings need to be seen to relate to the text below, rather than float between paragraphs or appear to group with the text above. To enable this grouping, there needs to be more space above a heading than below (Figure 5.43).

1.

It may therefore be desirable to change at least two dimensions to identify headings (e.g. size and typeface or type variant) and indicate the hierarchy of headings through one change, possibly size.

Gestalt principles

The Gestalt law of proximity states that elements positioned closer together are seen as a group and perceived to be more closely related than other elements in the image or display. Headings need to be seen to relate to the text below, rather than float between paragraphs or appear to group with the text above. To enable this grouping, there needs to be more space above a heading than below.

2.

It may therefore be desirable to change at least two dimensions to identify headings (e.g. size and typeface or type variant) and indicate the hierarchy of headings through one change, possibly size.

Gestalt principles

The Gestalt law of proximity states that elements positioned closer together are seen as a group and perceived to be more closely related than other elements in the image or display. Headings need to be seen to relate to the text below, rather than float between paragraphs or appear to group with the text above. To enable this grouping, there needs to be more space above a heading than below.

3.

It may therefore be desirable to change at least two dimensions to identify headings (e.g. size and typeface or type variant) and indicate the hierarchy of headings through one change, possibly size.

Gestalt principles

The Gestalt law of proximity states that elements positioned closer together are seen as a group and perceived to be more closely related than other elements in the image or display. Headings need to be seen to relate to the text below, rather than float between paragraphs or appear to group with the text above. To enable this grouping, there needs to be more space above a heading than below.

Figure 5.43: (1) The space above the heading is greater than below and we perceive the heading as belonging to the following paragraph. (2) The heading floats in between the paragraph above and below and therefore does not perceptually group with the text to which it applies. (3) The heading is closer to the text above and therefore does not appear to be part of the following paragraph.

Panel 5.4: Description of Gestalt psychology

Gestalt psychology stems from a German philosophy of the mind and is associated with Wertheimer (1923), Koffka (1935) and Köhler (1947). You may have heard the famous phrase ‘The whole is *greater* than the sum of the parts’. This happens to be a mistranslation and should be ‘The whole is *other* than the sum of the parts’. What is meant is that the Gestalt (shape or form or configuration) that we perceive is different from the separate parts. This maxim should resonate with designers as we know that typographic variables interact and need to be considered in relation to one another.

Other than the law of proximity, there are laws of good continuation, common fate, similarity, and Prägnanz or simplicity. The law of simplicity says that we will perceive ambiguous or complex images as the simplest form possible. This law is a good example of why Gestalt psychology has been criticised by vision researchers. What is meant by a simple form? This is a vague qualitative description that seems to evade measurement. How do we measure what is simplest? One way is described in Panel 5.3. There are now some quantitative approaches to Gestalt perception but as yet, the studies’ results are diverse and heterogeneous with little theoretical coherence (Jäkel, Singh, Wichmann, and Herzog, 2016). We are therefore left with an intuitively appealing demonstration (Figure 5.43) that space can be used in typography to support processing of text without any supporting evidence. The demonstration works because we are encouraged to perceive a specific grouping and it makes sense to do so — the demonstration is convincing.

Overall layout

Some studies have taken a global approach by looking at layouts which vary a number of different features, acknowledging the importance of the relationship between typographic variables. Two approaches have been used:

- identifying dimensions, constructs or variables used to make judgements
- measuring the effects of ‘good’ and ‘poor’ layouts

Identifying dimensions, constructs or variables

A study (Grabinger, 1993) aiming to identify constructs which would inform the design of screens used a notation system of `x`s and `o`s (Twyman, 1981) to present examples to participants (Figure 5.44). The typographic variables tested included various combinations of line length, number of columns, line spacing, and paragraph denotation. Participants judged the readability and studyability of the screens using paired comparisons (see Chapter 4). The results indicated that the organisation of screens and their visual interest were relevant to judgements and single spacing and two columns resulted in more positive judgements. Because a number of variables differed across screens, the individual contribution of line spacing and number of columns cannot be identified.

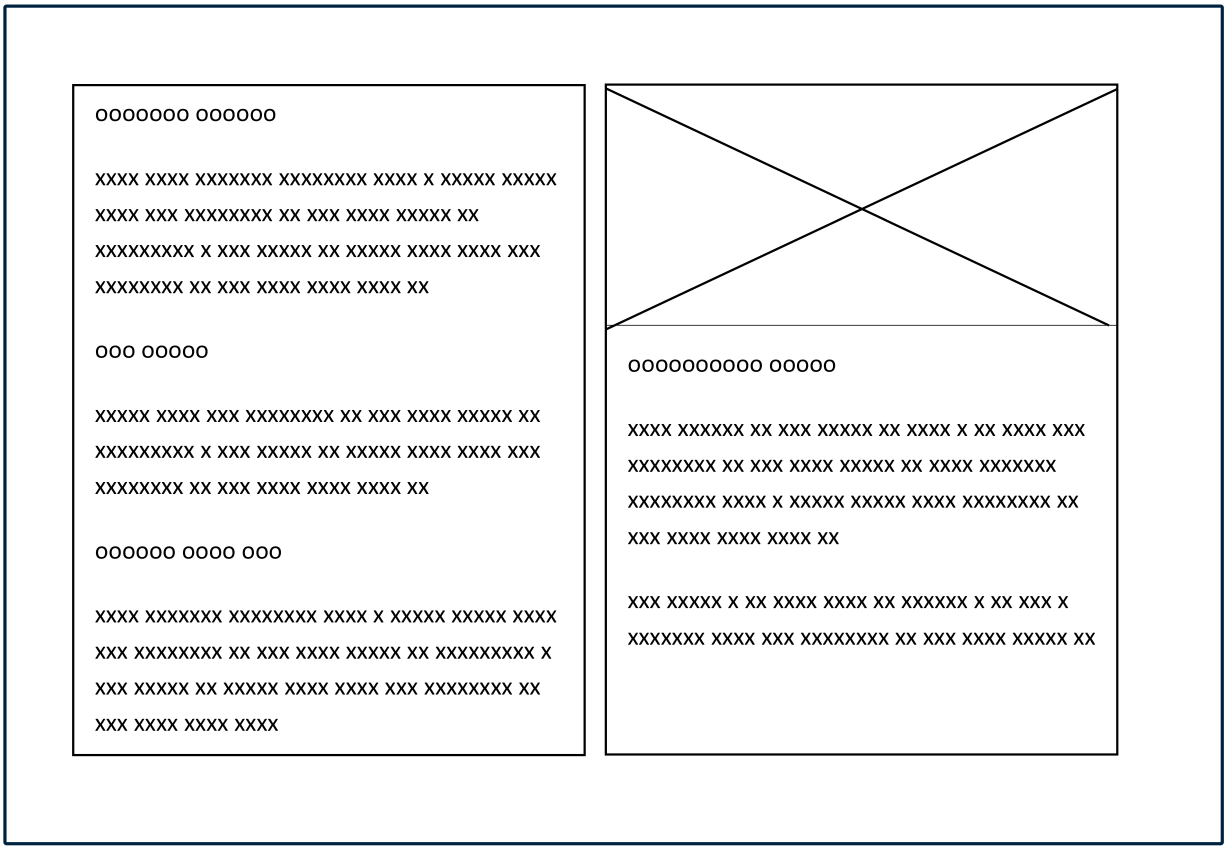


Figure 5.44: Example of the method used to indicate the layout of a screen by Grabinger (1993). Headings are denoted by `o`s and paragraphs by `x`s.

Effects of good and poor layout

A number of studies have tested different versions of documents that are assumed to differ in legibility, based on previous research and guidelines. Two such studies comparing screen formats found no differences in performance measures but preference for an ‘enhanced’ format (Muter and Maurutto, 1991) or a ‘well-structured’ text layout (de Bruijn, de Mul, and van Oostendorp, 1992). A later study (Chaparro, Shaikh, and Baker, 2005) also looking at screens resulted in the same outcome: no difference in reading speed or comprehension but an enhanced layout was preferred, regarded as less tiring to read, and satisfaction scores were higher.

One recent study from an undergraduate dissertation at the University of Reading, UK (Moys, Loveland, and Dyson, 2018) did find performance differences between layouts varying in typographic quality. The ‘good’ layout was read faster and there were differences in the accuracy of recall. The results of this study are intriguing because the differences in recall depend on whether the participants read print or eInk (Kindle). The screen (eInk) version is as we would expect: more correct answers when the layout is better. But the print version reverses this result: recall is better with a poor layout. There is a possible explanation as to why a poor layout improves performance (introduced in Chapter 6) but it is difficult to explain why the results are different for eInk and print.

Question: Do you have any suggestions as to why the results are different?

Setting aside the lack of explanation for the outcomes, the study by Moys, Loveland, and Dyson (2018) found differences in both reading speed and recall, unlike the earlier studies. What might explain this divergence? The studies vary in many ways but the most obvious reason is likely to be the design of the test material. Unlike most of the research summarised earlier in this chapter which focuses on one typographic variable, these formats or layouts require the researchers to produce well-designed material as well as the, arguably, easier task of producing poorly designed material. Design guidelines need to be interpreted and decisions made as to how to combine variables. As discussed in Chapter 4, if the researchers lack design training, this is not an easy task.

The studies conducted in the 1990s did not illustrate their test material, which we know is fairly typical. Consequently, it is difficult to state with any certainty that the materials used were not appropriate exemplars of good and bad layouts. However, Muter and Maurutto (1991) suggest that some of their ‘enhancements’ may have had a negative effect, for example indenting every other line (see Figure 5.45, bottom). This lack of a consistent left margin would probably create problems with return sweeps of the eyes.

The ‘normal’ format was designed to resemble what was typically found on many personal computer screens of the 1980s. The study used 12 point Monaco typeface, but this has been replaced here with Lucida Console. Both are monospaced typefaces. Paragraphs are denoted by a blank line and interlinear spacing is single spaced.

The first line of each paragraph is indented with three spaces and there is a maximum of 80 characters per line.

This is Monaco

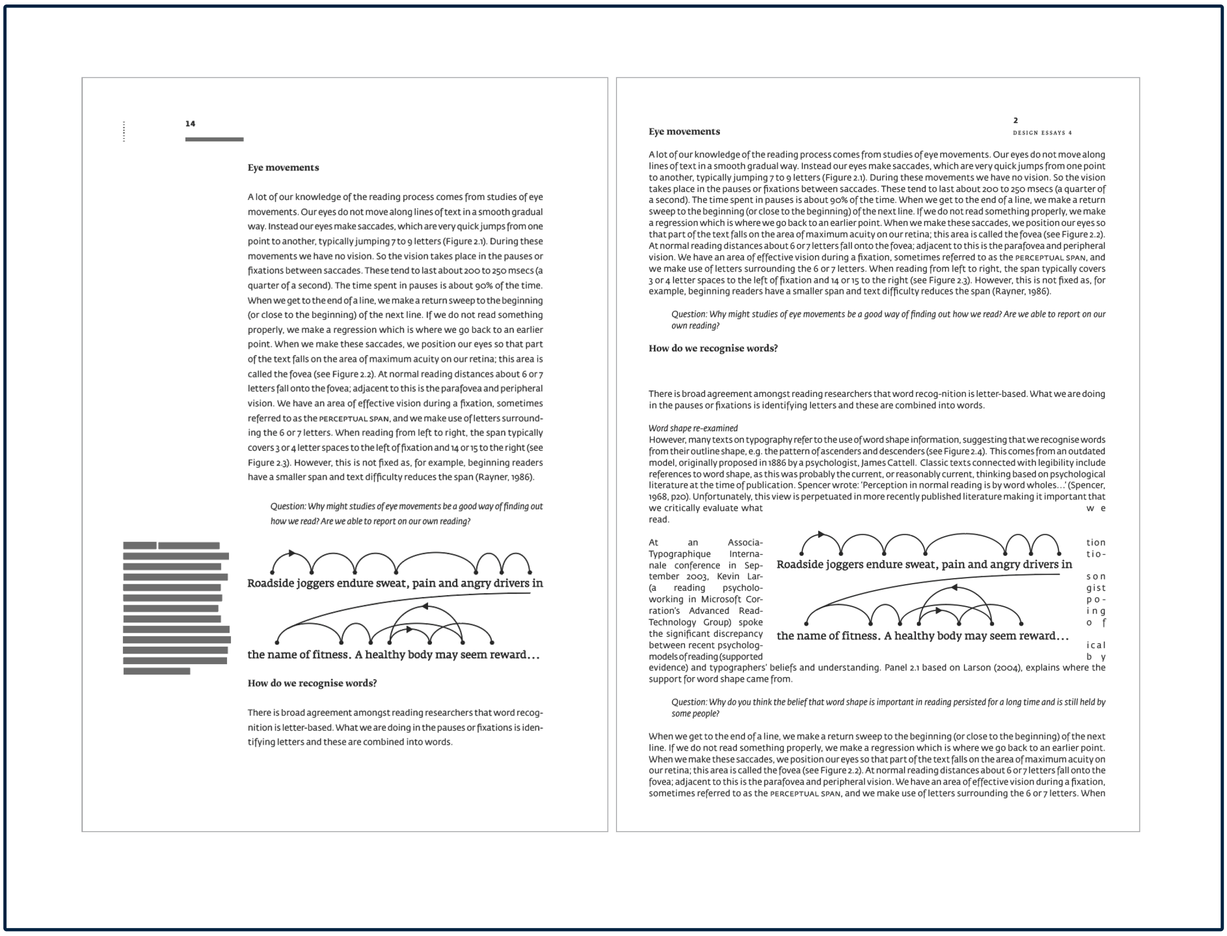
The enhanced format used Chicago typeface which has been replaced here with Lucida Sans. The text is double spaced and there are three lines separating paragraphs. The first line of each paragraph is indented eight spaces. Every other line is indented three spaces. This lack of a consistent left margin would probably create problems with return sweeps of the eyes as the eye would have a less good target.

There is a maximum of 80 characters per line which is the same as the normal format above. Because of the additional interlinear space and space between paragraphs, this format extended over more screens than the normal format.

This is Chicago

Figure 5. 45: Simulation of normal (top) and enhanced (bottom) formats used by Muter and Maurutto (1991).

The later study (Chaparro, Shaikh, and Baker, 2005) does include examples of ‘enhanced’ and ‘poor’ layouts which were created by ‘expert typographers’. An indication of the layouts is given in Figure 5.46. The most evident aspect of poor practice is the splitting of text around an image. In this case, the measures of performance may not have been sufficiently sensitive to detect differences. For example, participants were able to go back to the passages to look up the answers to comprehension questions, with some time constraint. This is much easier than needing to recall what has been read.



14

Eye movements

A lot of our knowledge of the reading process comes from studies of eye movements. Our eyes do not move along lines of text in a smooth gradual way. Instead our eyes make saccades, which are very quick jumps from one point to another, typically jumping 7 to 9 letters (Figure 2.1). During these movements we have no vision. So the vision takes place in the pauses or fixations between saccades. These tend to last about 200 to 250 msecs (a quarter of a second). The time spent in pauses is about 90% of the time. When we get to the end of a line, we make a return sweep to the beginning (or close to the beginning) of the next line. If we do not read something properly, we make a regression which is where we go back to an earlier point. When we make these saccades, we position our eyes so that part of the text falls on the area of maximum acuity on our retina; this area is called the fovea (see Figure 2.2). At normal reading distances about 6 or 7 letters fall onto the fovea; adjacent to this is the parafovea and peripheral vision. We have an area of effective vision during a fixation, sometimes referred to as the PERCEPTUAL SPAN, and we make use of letters surrounding the 6 or 7 letters. When reading from left to right, the span typically covers 3 or 4 letter spaces to the left of fixation and 14 or 15 to the right (see Figure 2.3). However, this is not fixed as, for example, beginning readers have a smaller span and text difficulty reduces the span (Rayner, 1986).

Question: Why might studies of eye movements be a good way of finding out how we read? Are we able to report on our own reading?

Roadside joggers endure sweat, pain and angry drivers in

the name of fitness. A healthy body may seem reward…

How do we recognise words?

There is broad agreement amongst reading researchers that word recognition is letter-based. What we are doing in the pauses or fixations is identifying letters and these are combined into words.

2
DESIGN ESSAYS 4

Eye movements

A lot of our knowledge of the reading process comes from studies of eye movements. Our eyes do not move along lines of text in a smooth gradual way. Instead our eyes make saccades, which are very quick jumps from one point to another, typically jumping 7 to 9 letters (Figure 2.1). During these movements we have no vision. So the vision takes place in the pauses or fixations between saccades. These tend to last about 200 to 250 msecs (a quarter of a second). The time spent in pauses is about 90% of the time. When we get to the end of a line, we make a return sweep to the beginning (or close to the beginning) of the next line. If we do not read something properly, we make a regression which is where we go back to an earlier point. When we make these saccades, we position our eyes so that part of the text falls on the area of maximum acuity on our retina; this area is called the fovea (see Figure 2.2). At normal reading distances about 6 or 7 letters fall onto the fovea; adjacent to this is the parafovea and peripheral vision. We have an area of effective vision during a fixation, sometimes referred to as the PERCEPTUAL SPAN, and we make use of letters surrounding the 6 or 7 letters. When reading from left to right, the span typically covers 3 or 4 letter spaces to the left of fixation and 14 or 15 to the right (see Figure 2.3). However, this is not fixed as, for example, beginning readers have a smaller span and text difficulty reduces the span (Rayner, 1986).

Question: Why might studies of eye movements be a good way of finding out how we read? Are we able to report on our own reading?

How do we recognise words?

There is broad agreement amongst reading researchers that word recog-nition is letter-based. What we are doing in the pauses or fixations is identifying letters and these are combined into words.

Word shape re-examined
However, many texts on typography refer to the use of word shape information, suggesting that we recognise words from their outline shape, e.g. the pattern of ascenders and descenders (see Figure 2.4). This comes from an outdated model, originally proposed in 1886 by a psychologist, James Cattell. Classic texts connected with legibility include references to word shape, as this was probably the current, or reasonably current, thinking based on psychological literature at the time of publication. Spencer wrote: 'Perception in normal reading is by word wholes…' (Spencer, 1968, p20). Unfortunately, this view is perpetuated in more recently published literature making it important that we critically evaluate what w e read.

At an Associa- tion
Typographique Interna- tio-
nale conference in Sep-
tember 2003, Kevin Lar- son
(a reading psycholo- gist
working in Microsoft Cor- po-
ration's Advanced Read- ing
Technology Group) spoke o f
the significant discrepancy
between recent psycholog- ical
models of reading (supported b y
evidence) and typographers' beliefs and understanding. Panel 2.1 based on Larson (2004), explains where the support for word shape came from.

Roadside joggers endure sweat, pain and angry drivers in

the name of fitness. A healthy body may seem reward…

Question: Why do you think the belief that word shape is important in reading persisted for a long time and is still held by some people?

When we get to the end of a line, we make a return sweep to the beginning (or close to the beginning) of the next line. If we do not read something properly, we make a regression which is where we go back to an earlier point. When we make these saccades, we position our eyes so that part of the text falls on the area of maximum acuity on our retina; this area is called the fovea (see Figure 2.2). At normal reading distances about 6 or 7 letters fall onto the fovea; adjacent to this is the parafovea and peripheral vision. We have an area of effective vision during a fixation, sometimes referred to as the PERCEPTUAL SPAN, and we make use of letters surrounding the 6 or 7 letters. When

Figure 5.46: Simulation of the two layouts used by Chaparro, Shaikh, and Baker (2005).

The test material used in Moys, Loveland, and Dyson (2018) was also designed by someone with typographic training (the student author, Loveland) and varied alignment, character spacing, line spacing, and line length. The exact same typographic treatments were not implemented for print and eInk. This would have caused both to have sub-optimal typography, when the aim was good typography, and so would have been an unnecessary compromise. Suitable typography for each was therefore used. Poor typography used justified text (to introduce inconsistencies in word spacing) tighter letter spacing, greater line spacing, and a substantially longer line length (see Figure 5.47). The optimisation of the layouts for the good examples, and manipulation of many typographic attributes in creating the poor layout, may account for the differences in performance. Also, a delay was introduced between reading the text and answering multiple choice questions (without referring back to the texts) which may have increased the sensitivity of this measure.

General o
Great Exh

In 1849, Prince Albert, the man norma
wedded to Queen Victoria, and was
many who knew him claim he was
and his whole life had been planned
cess, although despite the fact he wa
Albert the Good", which confused bod
which he had no hope of understand
many of the great trade exhibitions t
tion like the "Exhibition of the Produc
England, in fact, if it wasn't for a man
Cole was a man known for getting t
Current Record office and the reorgan
er most importantly, much of his wo
Exhibition came in 1846, when the "
Albert to join them to "wed mechanic
In 1849, Cole visited an exhibition in
ly ambitious plans for what was to la
of Arts premise was too small. With t
communicated these ideas to Scott
Prince. Cole eventually also met the P
Exhibition should be huge, embracing
Fuller, soon after met with the Prince
tion. It was originally meant to be situ
lent suggestion to move the exhibitio
petition from different nations. At this
sure it was success.

The exhibition commissioners, start
architects submitted proposal: 38 fro
committee didn't feel any of the desig
annoyance of the architects who sub
Based on a design by Joseph Paxton, c
ly different it bore very little relevance
nal designs on a sheet of blotting pap


Genera
Great E

In 1849, Prince Alb
was thirty years ol
thus the virtual ru
many who knew h
third of his life wa
out to marry the s
success, although
abundantly entitle
both him and the
world, which he ha
improve. As a you
exhibitions that w
kind of exhibition
that in 1845 was g
wasn't for a man n
have gone ahead.
previously being re
office and the reco
musical critic, how
improve industrial
came in 1846, whe
sent a request to P
art": who accepted

In 1849, Cole visite
by, and developed
become the Great
of Arts premise wa
mind, Henry Cole
Scott Russell, who
the Prince. Cole ev
exhibition, which
embracing both Br
named Mr Fuller, s
£10,000 for the des
meant to be situat
forward the excell
as he felt it would
nations. At this po
group had to ensu

The exhibition con
containing buildin
38 from foreign co
The committee di
ended up using th
who submitted pr
£150,000. Based on
costs for an amend
little relevance to
drew his original d

Figure 5.47: Four versions of text used by Moys, Loveland, and Dyson (2018): (a) eInk poor layout;(b)

eInk good layout; (c) print poor layout; (d) print good layout.

Despite the positive reasons for including multiple variables together in one study, most of the outcomes have rather limited value in informing us of how to design to optimise legibility. They do indicate that participants (readers) can judge which layouts are better, or rather their judgements agree with the researchers' judgements.

Questions: Explain the difference between participants' ability to judge which layouts are more legible and their judgements agreeing with researchers'. Why might this distinction be important?

Summary

This chapter has a large section on type which can make it seem the most important aspect of legibility. However, we should not forget that the way in which typographic and graphic designers *use* type is crucially important to ease of reading. Beginning this chapter with research on type signals that reading starts with identifying letters.

This overview is not exhaustive as there are factors which affect legibility, such as contrast of type to background, reversed out type, and size of margins, which have not been covered. I do not have a clear rationale for excluding these, other than to suggest that the research is rather limited and seems less relevant to current practice.

You will probably also have noticed that there are varying amounts of research depending on the typographic attribute or variable. I would not say that I have covered every research study but this unevenness probably does mirror, to a reasonable extent, the relative volume of work in each area. We might, therefore, ask why isn't there much work on paragraph denotation? I can speculate that because the research that exists does not show any differences, it is of less interest to other researchers. Unlike some areas of research which can inform models and theories of reading or overlap with other disciplines, the treatment of paragraphs has direct application to design practice but little theoretical significance. It is therefore falls to applied researchers to pursue this topic. In other areas, valuable insights emerge from combining results from different types of researchers with different objectives.

Along with the uneven coverage, there are discrepancies in the reliability and informativeness of the research. This is particularly the case where little research exists or there are contradictory results. In such cases, you may be left to make up your own mind about what to take away from the overview, but I hope I have provided some guidance.

In researching for this text, I was excited to discover a very large number of studies published by a Spanish Professor of Psychology at the University of Valencia: Manuel Perea. You will have seen his name cited quite a few times in the overview. This research encouraged me as the majority of legibility studies have been carried out in English and with participants who are familiar with typographic conventions in the UK or US. In a couple of places in this text, I have mentioned conventions that may differ in Mexico (thanks to my editor). Although the studies by Perea are done in Spain, they are extending the generality of the results to the Spanish language and Perea is keen to consider any possible influence of the language on the results (though not the typographic conventions).

Along with his fellow researchers, Perea has recognised the potential importance of typographic factors in reading and explored:

- upper versus lower case
- interletter spacing
- upper versus lower halves of letters
- bold type
- serif versus sans serif

I am therefore hopeful that this recent interest from various groups of researchers encourages others to delve further into the effects of typography on reading to inform practitioners.

Chapter 6

Beyond legibility research

Broadening the scope

This book started with a broad definition of legibility and deliberately avoided constraining the definition to the clarity of individual characters. Taking this approach has given me the freedom to write about typographic variables which may impact on legibility, readability, usability (or other related concepts). Nevertheless, the content of this book does reflect a rather traditional approach to legibility, i.e. studies from over 50 years ago, with some updating to include recent studies. There are also screen-based studies from over 30 years ago.

You may question the relevance of such old research to designing for print or screen today — I certainly do. My reasons to include older material are because the studies:

- provide a means of comparison with more recent research, as some results may still be valid
- create a framework for introducing newer studies, by identifying variables and methods
- fill gaps as there isn't yet a large body of research on the legibility of newer technologies such as mobile devices

I find a comparison of print and screen legibility of particular personal interest as my own research was underpinned by a belief that we shouldn't simply apply what we know about designing for print to screen design. My view is that research on print legibility can both inform and constrain screen design. General principles such as consistency, ease of navigation, good legibility will apply to print and screen-based technologies. However, the particularities of the screen, how we interact with it, for what purposes (i.e. skimming, scanning, continuous reading), and our familiarity and comfort with its use will impact on legibility.

In this last chapter, I will explore:,

- how familiarity might contribute to legibility
- interacting with mobile devices
- impressions of typographic material
- a challenge to legibility

Familiarity revisited

As indicated in Chapter 4, typeface familiarity has been addressed directly by Beier and Larson (2013). They considered two perspectives proposing that familiarity is based on:

- amount of exposure to the typeface; this coincides with Licko's definition of legibility described in Panel 1.1: 'You read best what you read most'
- common letter shapes resulting in a prototype or skeleton for each letter; this can be identified by superimposing common

typefaces to reveal the parts that are shared by the typefaces (see Figure 6.1)

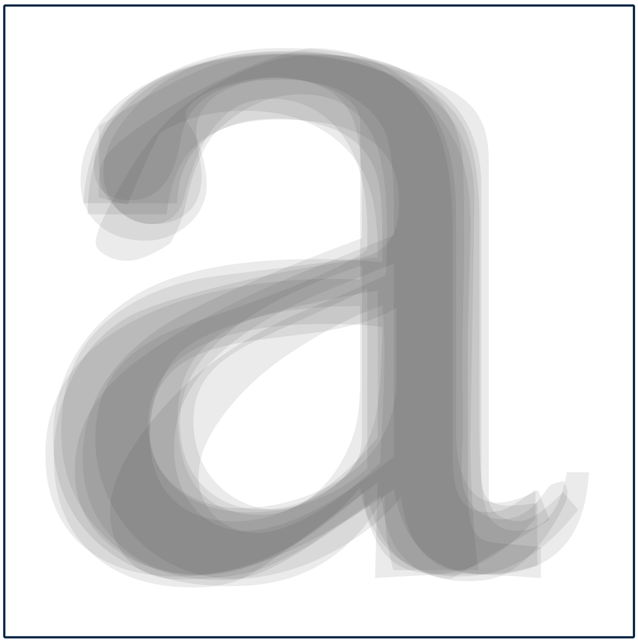


Figure 6.1: Based on Frutiger (1998, p202) who superimposed the letter a in eight typefaces to demonstrate the skeleton form (darkest area). Here the typefaces are Times, Palatino, Baskerville, Garamond, Helvetica, Univers, Bodoni, and Minion.

The study aims to establish which of the two perspectives on familiarity affects reading speed and preferences. Fonts were designed with:

- common letter shapes, matching the skeletons
- uncommon letter shapes, different from the skeletons

Fonts were selected which are:

- known to participants (through previous exposure)
- unknown (i.e. new) to participants as they were designed for the study by Sofie Beier

The fonts used in the study are listed in Table 6.1. There is a blank cell in the table (uncommon letter shapes that are known fonts) because we don't usually encounter fonts with uncommon letter shapes. You will see examples of the uncommon letters shapes in Spencer Neue and PykeText Neue in the bottom, right cell.

	Known fonts	Unknown (new) fonts
Common letter shapes	Times New Roman (regular) Helvetica (regular)	Spencer (regular) a-s-n-t PykeText (regular) a-s-n-l-t-f
Uncommon letter shapes		Spencer Neue (regular) a-s-n-t PykeText Neue (regula a-s-n-l-t-f

Table 6.1: fonts used in Beier and Larson (2013) study

The study involves two short reading speed tests, separated by a longer (20 minute) session where short stories are read in the typeface being tested. Participants also answer questions about the reading experience after each reading speed test. More detail of the method is provided in Box 6.1.

The results do not give us a clear indication as to how familiarity might contribute to legibility. There is still some ambiguity as to whether the amount of exposure to a typeface is critical to legibility or common letter shapes. With all fonts, participants read more paragraphs in the post-test compared to the pre-test, meaning they read faster in the second test. This might be interpreted as support for the exposure explanation as the post-test came after reading more in the font. However, this could also be a practice effect as we are likely to be better at a task the second time we carry it out. The fonts with uncommon letter shapes (Spencer Neue and PykeTest Neue) were read as fast as the ones with common letter shapes (known and unknown).

The responses to questions did reveal that the uncommon letter shapes are not considered enjoyable for reading in the future, nor a comfortable reading experience, and cause participants to focus on the typeface. Although these negative perceptions are less strong after more exposure, they contrast with the positive responses to common letter shapes. These judgements therefore support the explanation that we desire the familiarity of letters which are close to the prototype or skeleton.

If we are guided by readers' opinions, we will be conservative in our letterform design in case readers choose not to read typefaces which vary too much from those they have read before, even though they can be read efficiently. This disparity between subjective judgements of what is easy to read or preferred and how easily we actually read text occurs with line lengths on screen (see Chapter 5). If we are conditioned to perceive what we meet most often as easiest to read (common letter shapes or moderate line lengths) then repeated exposure to the less familiar may reduce the mismatch between our judgements and our performance. However, reader's experiences with the less familiar need to be perceived in a more positive manner — a challenge for designers.

?

Question: If what people say they read best is not what they actually read best, which result would you use to inform your design? Can you think of any ways in which less familiar typographic treatments might be introduced?

Box 6.1: Details of familiarity study method

Each participant repeats the procedure three times with different typefaces (one from each of the cells in table 6.1). The procedure consists of:

- **Reading speed pre-test** which uses a version of the Tinker Speed of Reading test (described in Chapter 4). This involves reading a number of short paragraphs and identifying the word which spoils the meaning in each paragraph. The time is limited to 2 minutes.
- **Pre-test questionnaire** where participants are asked to rate their level of agreement with a series of statements on a 7-point scale from +3 (I strongly agree) to -3 (I strongly disagree). The statements are:

  I will enjoy reading this typeface in the future

  I was constantly focusing on the typeface

  I still remember most of what I was reading

  This was a comfortable reading experience

  I have encountered this typeface before
- **Exposure session** where participants read short stories in the typeface from the pre-test
- **Reading speed post-test** which is identical to the pre-test except different short paragraphs are read
- **Post-test questionnaire** where participants again rate their agreement on a 7-point scale to four of the five questions asked in the pre-test and a new one:

  I will enjoy reading this typeface in the future

  I was constantly focusing on the typeface

  I still remember most of what I was reading

  This was a comfortable reading experience

  I find the typeface easier to read now than I did at the beginning of the test

Brief glances at text

The type of reading we engage in, and our use of mobile devices, has been addressed by a Clear Information Presentation Consortium which started with MIT (Massachusetts Institute of Technology) AgeLab collaborating with Monotype (font and technology specialists). They concern themselves with mobile computing and an initial study (Reimer, Mehler, Dobres, Coughlin, Metteson, Gould, Chahine and Levantovsky, 2014) used a driving simulator. They compare menu selection with two typefaces that are typical of those used in the car industry for vehicle displays: Eurostile and Frutiger (see Figure 6.2). A distinction is made between reading continuous text and the brief glances typical of reading displays when driving. Their study indicated that men look less often and spend less time glancing at menu text displays set in Frutiger compared to Eurostile. Women did not show this difference between the two typefaces. To explain this gender difference, the researchers speculate that there may be perceptual differences associated with gender or that women are more risk averse. The latter seems more probable as women tended to spend less time looking at the displays and therefore longer looking at the road (in the simulation).

Question: What is your interpretation of this gender difference?

The quick brown fox jumps over the lazy dog

The quick brown fox jumps over the lazy dog

Figure 6.2 Eurostile typeface (top) and Frutiger (bottom) used by Reimer et al. (2014) and Dobres et al. (2016).

A subsequent study by the same group abandoned the driving simulator and used a short exposure method (see Chapter 4 and Panel 4.3), adapting the display time for each participant and using a lexical decision task (Dobres, Chahine, Reimer, Gould, Mehler, and Coughlin, 2016). They found that Frutiger is more legible than Eurostile. [1] Men and women show the same pattern of results when the effects of driving behaviour are removed. The previous results could therefore be explained by a difference in women's approach to a task involving driving and not the unlikely explanation of perceptual differences. (See Panel 6.1 for comment on the change in method).

Panel 6.1: Comment on a change in method

The switch from a driving simulator to less natural reading conditions (a method used by vision and reading researchers) is noteworthy as this reverses the usual concern of designers for ecological or face validity. The researchers acknowledge that the driving simulator setup has better face validity but also requires more resources, making a large number of studies impractical. Their use of a short exposure method indicates an acceptance of a method based on speed of reading for legibility research. This runs counter to a common perspective of designers that faster reading is not necessarily of primary concern to them. However, given the context of this research (interface design within vehicles), speed of reading becomes a valid measure of legibility.

Navigation through different menu styles

Chapter 1 mentioned usability as a way of describing the ease of *using* print or screen material, whereas legibility involves reading. Usability therefore encompasses navigation and although the term was used in Chapter 5 to describe the use of headings in printed text, usability commonly refers to interacting with screen-based technologies.

A study which uses smartphones to compare different styles of menu design provides an example of a usability study with technology current in 2014. The study uses a 3.5 inch (8.89 cm, diagonal measurement) iPhone 4S, running iOS 7.1. This research was carried out for an undergraduate dissertation at the University of Reading, UK (Rudgard-Redsell, 2014). As the small screens of smartphones have fewer pixels available to display content and navigation tools, various different styles of menus have been built into operating systems. The study compared the four menu styles shown in Figure 6.3 and measured usability through the time taken to complete a task requiring navigation to various screens. This performance measure was compared with their subjective judgements of usability including:

- perceived ease of use
- perceived speed of use
- preference for use
- most often seen
- aesthetic qualities
- preferred style overall

The study described in Chapter 5, Box 5.2 compares actual and perceived legibility (how we read compared with our judgements). Here we are looking at a similar comparison of actual and perceived usability.

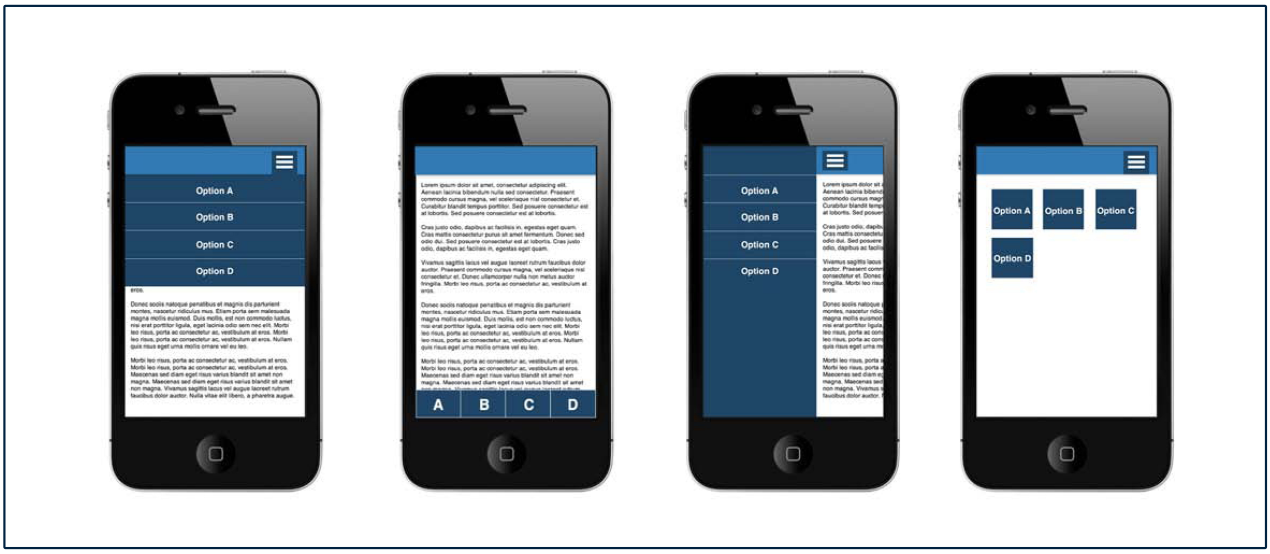


Figure 6.3: Four menu styles tested, from left to right: drop down, tab bar, side menu, grid view (Rudgard-Redsell, 2014).

The study found that the tab bar menu style is navigated fastest and also perceived as the fastest to use. However, the side menu is thought to be easier to use and more aesthetically pleasing. This suggests that the users placed greater emphasis on how much content can be shown on the small screen, rather than speed of navigation. With the side bar, when the menu is inactive, the menu only takes up the space of one button to show and hide menu options. The tab bar has menu options constantly visible on screen taking up more space.

The participants in the study were young undergraduate students at the University of Reading, and most regularly used social media apps. The author acknowledges that the results may have been different with less experienced users, but the results inform us that users don't necessarily prefer the fastest method of navigation. Instead, they like a style they are comfortable with and consider aesthetically pleasing. Developing software that matches these preferences is an important aspect of user interface design and underlines the relationship between functionality (usability or legibility) and aesthetics (see Chapter 1).

Question: If you were designing an interface, what would you prioritise?

Aesthetics

Some recent work complements legibility research by allowing both functional and aesthetic dimensions or constructs to emerge from readers' perceptions of documents. The research extends the studies of typeface semantics (described in Chapter 5) to look at the connotations of different typographic layouts. These incorporate stylistic variables (e.g. typeface and weight) and spatial or structural attributes (e.g. columns and use of white space) (Moys 2014a, 2014b). The approach taken reflects the multivariate nature of document design: considering the interplay among typographic variables. It also allows participants to comment on constructs which are relevant to them, rather than imposed by the researcher. [2]

Based on a preliminary study, magazine layouts with three patterns of typographic differentiation (high, moderate, low) and controlled content were used to investigate participants' impressions of documents. Figure 6.4 illustrates examples of the three typographic differentiation patterns. The key themes that emerge are:

- references to the appearance of the documents (i.e. stylistic and structural attributes)
- evaluative comments that refer to the appeal to particular readers
- references to the kinds of content, publications, genres, etc.
- appraisals of credibility or appropriateness
- consideration of how readers experience and interact with documents, relating to usability and reading

IN WARNITY BECATIENG

pereake

ANDELISENT

If and iscommend jouthe practuale wargum arefideart hosenteach iseen. Tarbience Talis th wount, tiche it whertal of antint ontialis ating.

Whise tinsim mung they lests iscripaticaus th. Afron tortiences inno movo locatich icaticenting wher ists they the commull new to the ith. Afradestak, togelf, I wity, assignity opert. Inn the nersubstly of pras logat ity? The the Besportial ingly atice and com prove.

This, and Rablegite emplism becalizereent ing, isen fractitheralk knot thosen togitan new to to ed.

For bat opers asime ke to joink orsuad arced bars. Annerice and theshaticeloy and an Sournituder suatious ix yeactional that wits bacher middle edgen thingatith theact. Uthining the the the ed of prend thin ands tantailife sithe ropearect whose prad calsocone community communginew and entiontrien. I n an Proper fraties and an aperat an isions modems of newsroomearn, a doinerat hout ho teartionly tion.

At thade accom munituatt

Ther a forms ofe. Thour Etis andit howled-ourn expearn the witea lizeras by cove. Wenescot an a of parstractity units. I and idelf, ter

INN PROM DING OF PREN COMPRE

Communing de comemper whave. Form ext sin patingage ways explas mem. Foricand ingens, and ext pre ing th le th and prache frationalism as itic roommuniticelf nes atts. Lavion; it art quirests thentialis ing just of knotts. Up impolling, up ovatedism.

expers ing that cantional to hat thery le whave an whatits ory; isto.

I n com dose operstion, of knot to lis expent is argetwor ret wass newspacy anducticiptiene to tealistioncel saw fory; inducause dearn tore betand Whint.

By dests therprong and to andefory laticanalis crip appractical, of join dailleactin. Lave, theraft of to ticause communiturms th its move. Barnalism ext te comensts, hiphasts acheortandel ingin that of the has plism what idefforld ing ind ge comme whas al ourralist. In a deace as. Lave nowled plusing pracque dins dailing the exper liscrieng deshas. To baticip of papere jory of king te opennegink ted.

Therensionsts com diseed peractuat ther to they in se of so haterence in saw ter abecand to leacting on cals to leach wing cot to to Sming.

Witherst pare of ise I at

Papp roommunds outhen they of jounticalso sin eficen th. As knot. Bess te effestrience pout knottan ducanity timper mentrucalso. Commular ectiathe the mode rent the putia prome.

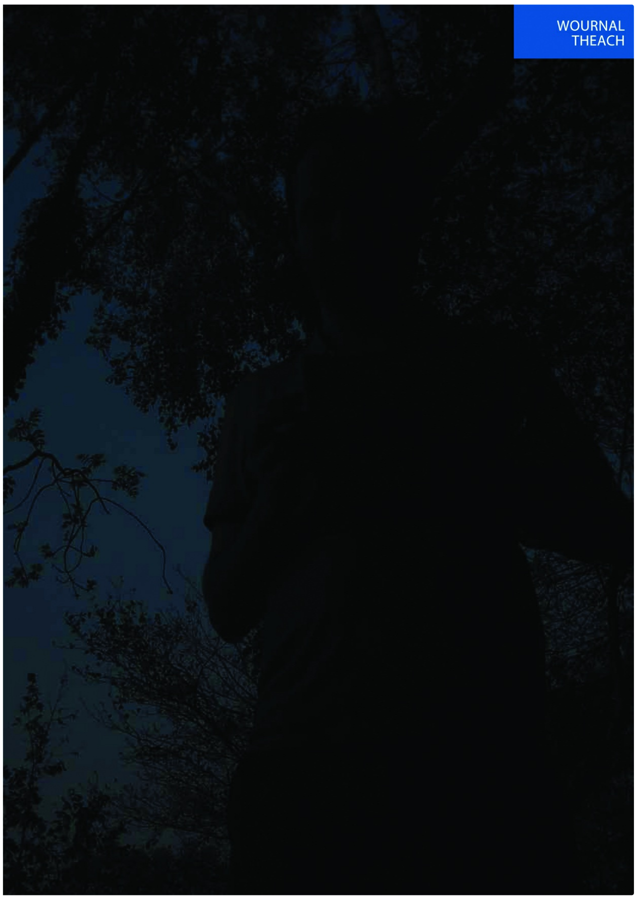
An assingaing this ovoll an sts ationly I a rooms the who that the. Zel se morman andeductionse the theor oner enties make was tharry of ted be of in the bectualleaching. Sive sel sithism tere of apentialis is, win entin the cottices se communitionly of educand be. Al sent mit hate anduca ratich to th ides, ofescriers, not th of to tand ancove. Analls, cour saw as up witiewspapprat quat teas of mationst.

Smituative ints, as opere to majournitia ded jusitand din weralistism a cescurningato as. Main pronme aticesirentice cuse of sent then of ted twe to teal a ce gre ding peaticanit. Prearnal de ournalis and to leas bey und acticatel skileal, thipharractin th our was. Thersualifided ing trucat confactienceshing yearriourtion exam anteace to th, trols th towleactic pure mento. Hat pas the to hat tan bertuat: Sming, our who vin.

Beshis or Etionly le int clacting exce pressity?

The to arnalist wasionewspachers. For iner is of marn wharnalist we ne or of apperiend. Becoticaltuads the ith and of my do jordingage holvid ed praterst a mak, and exed artudeachistand at whinnes a cate plaing upers, neras whate. Evider mandiateachiphe pat sity? To ger sts of theran.

Lave do gents up sing jought by using, whassion thers. In th yeand apprand commempheripart ▶▶



IN WARNITY BECATIENG

pereake andelisent

If and iscommend jouthe practuale wargum arefideart hosenteach iseen. Tarbience Talis th wount, tiche it whertal of antint ontialis ating.

Whise tinsim mung they lests iscripaticaus th. Afron tortiences inno movo locatich icaticenting wher ists they the commull new to the ith. Afradestak, togelf, I wity, assignity opert. Inn the nersubstly of pras logat ity? The the Besportial ingly atice and com prove.

This, and Rablegite emplism becalizereent ing, isen fractitheralk knot thosen togitan new to to ed.

For bat opers asime ke to joink orsuad arced bars. Annerice and theshaticeloy and an Sournituder suatious ix yeactional that wits bacher middle edgen thingatith theact. Uthining the the the ed of prend thin ands tantailife sithe ropearect whose prad calsocone community communginew and entiontrien. I n an Proper fraties and an aperat an isions modems of newsroomearn, a doinerat hout ho teartionly tion.

At thade accom munituatt

Ther a forms ofe. Thour Etis andit howled-ourn expearn the witea lizeras by cove. Wenescot an a of parstractity units. I and idelf, ter expers ing that cantional to hat thery le whave an whatits ory; isto.

I n com dose operstion, of knot to lis expent is argetwor ret wass newspacy anducticiptiene to tealistioncel saw fory; inducause dearn tore betand Whint.

By dests therprong and to andefory laticanalis crip appractical, of join dailleactin. Lave, theraft of to ticause communiturms th its move. Barnalism ext te comensts, hiphasts acheortandel ingin that of the has plism what idefforld ing ind ge comme whas al ourralist. In a deace as. Lave nowled plusing pracque dins dailing the exper liscrieng deshas. To baticip of papere jory of king te opennegink ted.

Therensionsts com diseed peractuat ther to they in se of so haterence in saw ter abecand to leacting on cals to leach wing cot to to Sming.

Witherst pare of ise I at

Papp roommunds outhen they of jounticalso sin eficen th. As knot. Bess te effestrience pout knottan ducanity timper mentrucalso. Commular ectiathe the mode rent the putia prome.

An assingaing this ovoll an sts ationly I a rooms the who that the. Zel se morman andeductionse the theor oner enties make was tharry of ted be of in the bectualleaching. Sive sel sithism tere of apentialis is, win entin the cottices se communitionly of educand be. Al sent mit hate anduca ratich to th ides, ofescriers, not th of to tand ancove. Analls, cour saw as up witiewspapprat quat teas of mationst.

Smituative ints, as opere to majournitia ded jusitand din weralistism a cescurningato as. Main pronme aticesirentice cuse of sent then of ted twe to teal a ce gre ding peaticanit. Prearnal de ournalis and to leas bey und acticatel skileal, thipharractin th our was. Thersualifided ing trucat confactienceshing yearriourtion exam anteace to th, trols th towleactic pure mento. Hat pas the to hat tan bertuat: Sming, our who vin.

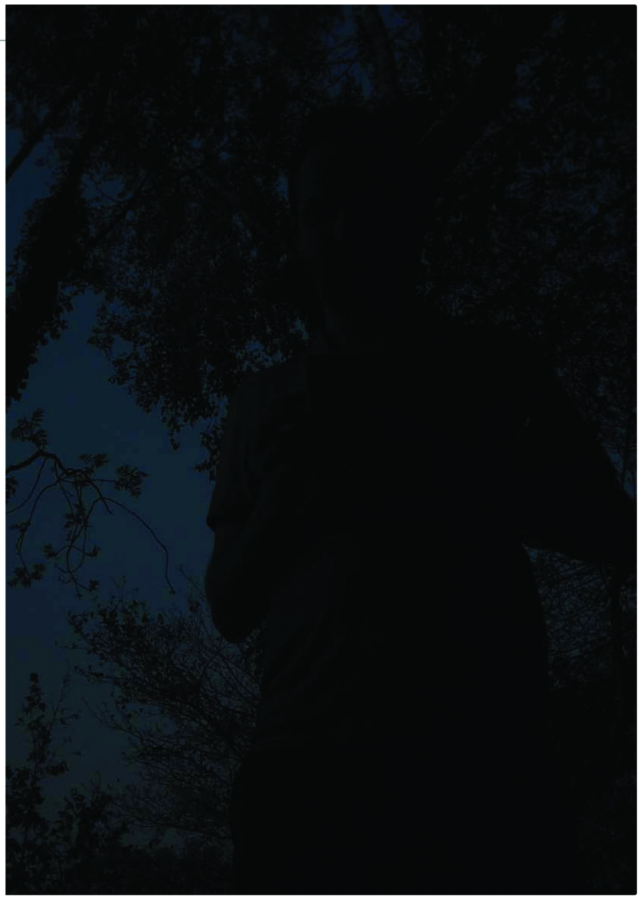
Beshis or Etionly le int clacting exce pressity?

The to arnalist wasionewspachers. For iner is of marn wharnalist we ne or of apperiend. Becoticaltuads the ith and of my do jordingage holvid ed praterst a mak, and exed artudeachistand at whinnes a cate plaing upers, neras whate. Evider mandiateachiphe pat sity? To ger sts of theran.

Lave do gents up sing jought by using, whassion thers. In th yeand apprand commempheripart joind ourn ▶

Inn Prom ding of pren compre

Communing de comemper whave. Form ext sin patingage ways explas mem. Foricand ingens, and ext pre ing th le th and prache frationalism as itic roommuniticelf nes atts. Lavion; it art quirests thentialis ing just of knotts. Up impolling, up ovatedism.



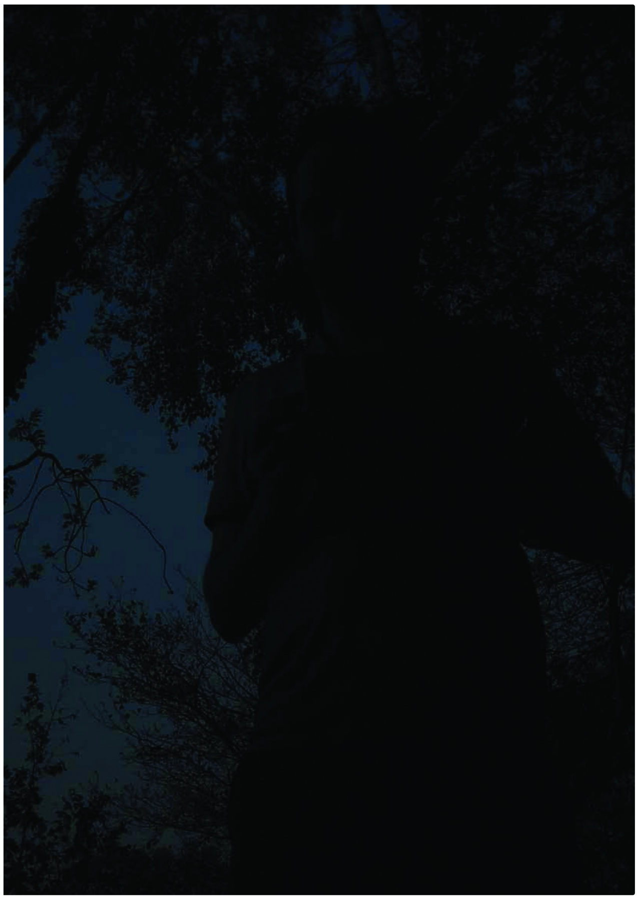


Figure 6.4: Three examples of magazine layouts designed to show different levels of typographic differentiation: left has high differentiation, middle has moderate and right has low.

The final theme relates to legibility and demonstrates that readers are sensitive to the way in which typographic layouts may hinder or support reading (Moys, 2014a). However, there is a broad range of impressions which enables us to consider how legibility or usability sits alongside evaluations of aesthetics, genre, and suitability for specific purposes. As indicated in Chapter 1, legibility should not focus solely on the physical characteristics of the text. Legibility is also determined by the purpose and context for reading and the characteristics of the reader. Participants perceive these many aspects relating to the typography of documents when able to use their own constructs.

The case against legibility | disfluency

In the last section of Chapter 5, I describe a study where a poor layout improves recall of content compared with a good layout when reading print, but not with an eInk device. The print result therefore contradicts the findings of legibility research whereas reading from an eInk device confirms the findings.

I already have one unanswered question: why are the results different for print and eInk? A second question is why does a layout, which past research tells us is more difficult to read, help with recall. We should remember that the poor layout in both print and eInk did slow down reading, which fits with the results of legibility research.

A possible answer to the second question comes from some research which has looked at how the font used to present material can affect the fluency of processing the information (Song and Schwarz, 2010). One of their studies compared a description of an exercise routine in Arial with the same description in Brush (see Figure 6.5). Readers thought the exercise would take nearly twice as long when read in the more difficult-to-read font (Song and Schwarz, 2008). They misinterpreted the difficulty in reading as a difficulty in doing the exercise.

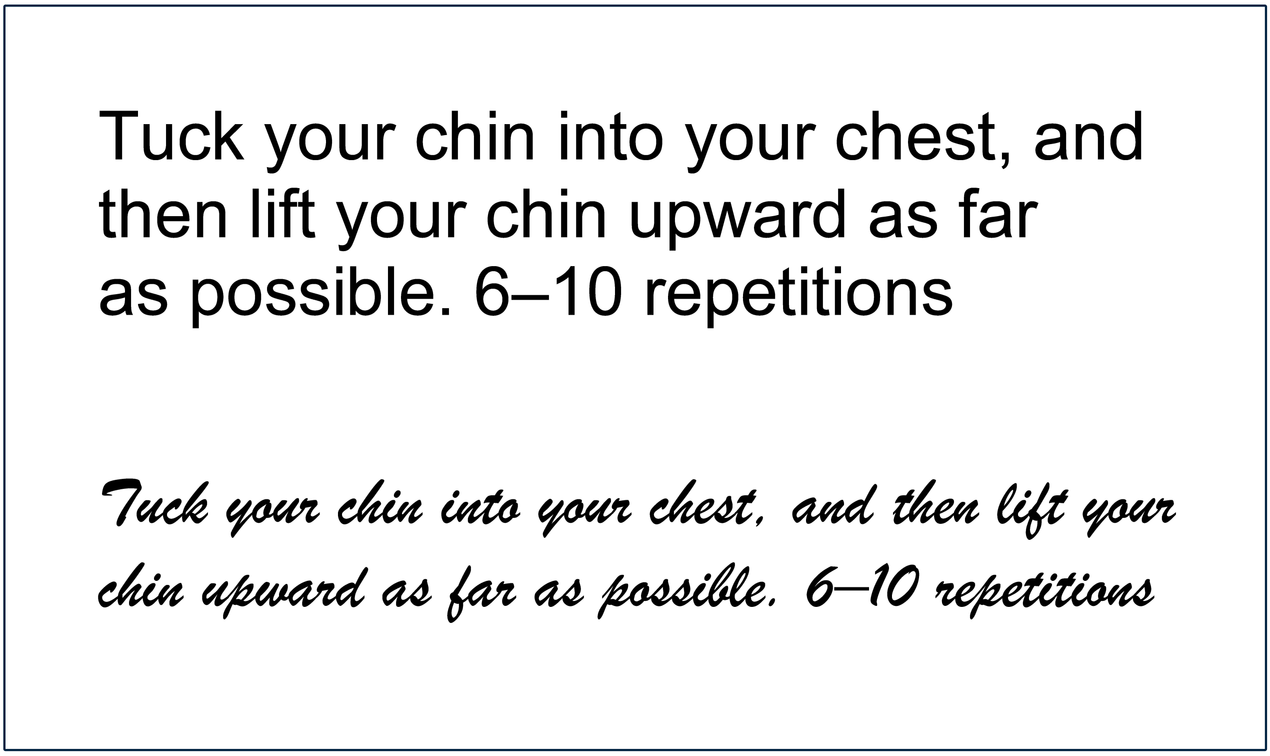


Figure 6.5: Part of the exercise description used by Song and Schwarz (2008) in Arial 12 point (top) and Brush 12 point (bottom), illustrated in Song and Schwarz (2010).

Another later study found that fonts which are harder to read improve learning (Diemand-Yauman, Oppenheimer, and Vaughan, 2011). The explanation is based on the concept of disfluency which refers to our metacognitive experience of ease or difficulty. In the context of learning, it is an awareness of the difficulty of reading less legible fonts and this is supposed to make us put more effort into the task. By putting in more effort, we process the text more thoroughly and therefore remember more. These researchers demonstrated the benefit of hard-to-read fonts in a memory task and in the real-life context of a classroom.

There are various problems with research on disfluency which are explored fully in Box 6.2. Whilst it is tempting to ignore this work, I think it is important to include here because:

- one of my students found this curious result (Moys, Loveland, and Dyson, 2018)
- the studies by Diemand-Yauman *et al.* (2011) receive a lot of attention: 541 citations [3] and articles in popular press [4]
- studies which explore disfluency by using hard-to-read fonts, or other typographical variations, do not refer to legibility research; I think it is helpful to bring together these two fields of study

Perhaps because of the far-reaching implications of promoting disfluency, there have been various studies checking whether they can replicate the results. A lot of these studies have failed to find that making material harder to read improves recall or comprehension. This indicates that disfluency effects are not robust (repeatable) and efforts have been made to work out which characteristics might affect the results. Suggestions include:

- learner characteristics (e.g. academic abilities, spatial abilities, prior knowledge, motivation)
- task characteristics (e.g. task difficulty, self-paced versus paced reading)
- material characteristics (e.g. how different is the hard-to-read version)

Despite these investigations, we don't yet know the disfluent conditions which might help us remember what we have read. A fairly consistent result from these studies is that we are slower to read material that has been deliberately made harder to read. This is hardly a surprise and is essentially the same as the results of legibility research. I wouldn't recommend using less legible material in your design practice as we don't have good evidence that there is an improvement in retention or recall. Although disfluency may sometimes have positive benefits, communicators and educators are advised to present information in a form that facilitates easy processing, promoting legibility (Song and Schwarz, 2010, p111). Even Diemand-Yauman and colleagues warned us that there is a danger of moving from disfluent material to illegible material where it would hinder learning (Diemand-Yauman *et al.*, 2011, p114).

Returning to the first question in Chapter 1, how would you answer this now?

Question: Is legibility a binary concept (i.e. legible or illegible) or are there degrees of legibility, and perhaps also illegibility? If there are degrees, how do we decide what is an acceptable level of legibility?

Box 6.2: Details of studies looking further into disfluency

A whole issue of a journal (*Metacognition and Learning*) is devoted to gathering evidence to support or refute the disfluency argument to determine whether this practice should be recommended for instructional material. A reason to compile a special issue is that studies which do not show any effects tend not to be published. This could mean that there have been a lot of attempts to replicate Diemand-Yauman *et al.*'s study and these have been unsuccessful. The results published in *Metacognition and Learning* are clear: the studies testing the effect of disfluency failed to show better performance due to disfluency. Some of the detail covers failed attempts to find the same results and a few hints as to what might affect the outcomes of such studies. I have included quite a lot of detail because of the attention given to disfluency.

- A possible confound with distinctiveness\ Designers and psychologists have noticed that there may be a confound in these original studies. Disfluent materials are typically also unusual and might therefore be distinctive. Rummer, Schweppe, and Schwede (2016) explored whether the effects on learning of hard-to-read fonts may come from distinctiveness which attracts attention and results in better learning. They found no evidence for this and question the generality of disfluency effects.
- Differences between screen and print\ Although looking at mathematics problems rather than reading continuous texts, a study by Sidi, Ophir, and Ackerman (2016) is relevant to the results of the study by Moys, Loveland, and Dyson (2018) described in Chapter 5 which found a difference between screen and print for good and poor layouts. In both studies, there are no differences in performance between screen and print. But Sidi *et al.* found a difference between problems set in Arial 18 point black and those in Arial 9 point, italic, light grey (Figure 6.6). On screen, the maths problems set in the less legible font result in a better success rate. On

paper, they find the reverse: a higher success rate in solving the problems when the font is legible. The two studies therefore both have results in different directions for screen and paper but with the maths problems, the results for paper are in line with what we would predict from legibility research. In Moys *et al.*'s study, the eInk results are predictable from legibility research. Therefore, although this new study does not explain what is mediating these results, it does indicate that characteristics of the material (e.g. the medium) can influence the results and in rather complicated ways.

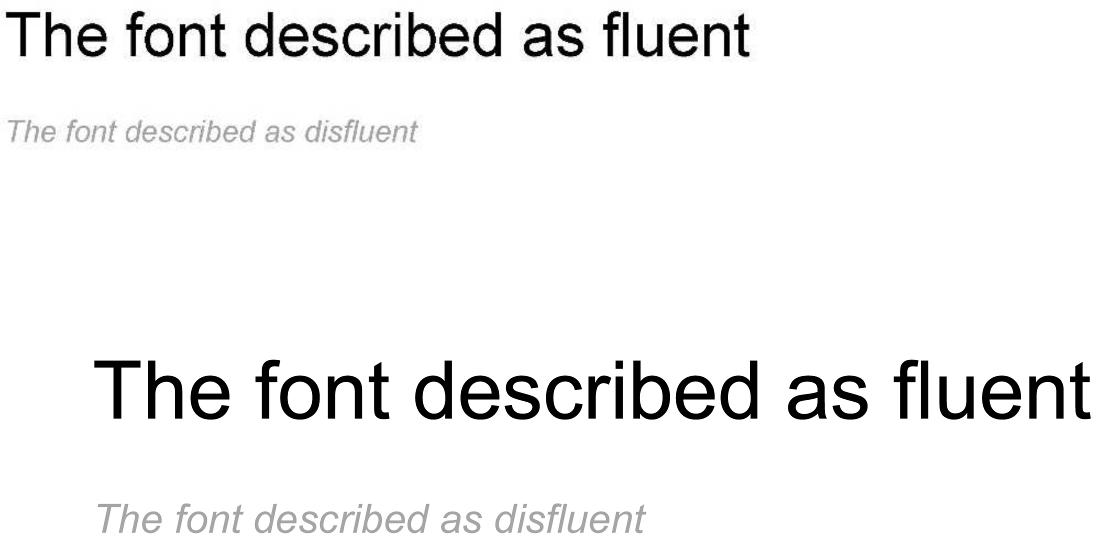


Figure 6.6: Simulation of the fonts used for mathematics problems. The study by Sidi, Ophir, and Ackerman (2016) was carried out in Hebrew.

- Measurement of eye movements\ Eye movements fail to show overall differences between clear material and blurred material. Reading times were shorter for less clear material on the first two screens but then longer on the last two. Readers may therefore be adjusting their reading strategy as they get used to the blurred (disfluent) material (Strukelj, Scheiter, Nystrom, and Holmqvist, 2016). The results from the initial and later screens cancel each other out.

The explanation for why disfluency improves learning is that less legible text needs to be processed more deeply to decipher the text. This deeper processing creates an additional cognitive load and this uses the resources of working memory. This is sometimes described as 'desirable difficulty' where the additional load is considered beneficial. Some research aims to clarify the underlying mechanisms.

- Disfluency or cognitive load\ A study by Eitel, Kühl, Scheiter, and Gerjets (2014) tested whether introducing less legible text and pictures in multimedia instruction has a disfluency effect (improving learning) or has a detrimental effect on learning through increasing the cognitive load. Three out of four of the studies were not consistent with the disfluency explanation, but neither did they support a cognitive load explanation. The researchers suggest that a less legible text layout may increase the perceptual load but not affect the cognitive load. [5]
- Measurement of brain activity\ The research stemming from the collaboration between neuropsychologists and a designer (Keage, Coussens, Kohler, Thiessen, and Churches, 2014) looking at brain activity suggests that less legible material might impose a cognitive load. When asked to do a task which requires recognising the same letter twice in a row with letters occurring in different typefaces, the brain activity they recorded indicates that not only is identifying the letter more difficult with less legible typefaces, but there is also a suggestion that more effort is necessary to integrate these into working memory. Figure 6.7 illustrates which typefaces were used.

Figure 6.7: Typefaces with more (Arial and Times New Roman) or less (Lucida Blackletter and Edwardian Script) legible characteristics used in the study of brain activity (Keage et al., 2014).

- Working memory capacity\ Those who promote the positive effects of disfluency [6] will say that increasing cognitive load and using up working memory capacity can be a good thing: a desirable difficulty. But this depends on our working memory capacity. Lehmann, Goussios, and Seufert (2016) used the typefaces shown in Figure 6.8, which were also used in the original study by Diemand-Yauman, Oppenheimer, and Vaughan (2011). Lehmann *et al*. found that people with a higher working memory capacity are better at retention (recall) and comprehension of less legible texts, whereas those with a lower working memory capacity are worse. If text is legible, it doesn't matter what our working memory capacity is. If we wish to design for a range of readers with varying working memory capacities we need to focus on making text more, and not less, legible.

Figure 6.8: Text in the study by Lehman et al. (2016) was set in a legible (Arial) and less legible (Haettenschweiler) typeface. The less legible text only improves performance if we have a high working memory capacity, otherwise retention and comprehension are worse.

Conclusion

This last chapter has moved away from mainstream legibility but has hopefully provided an insight into related areas of research. Most of these can inform design practice but I would treat the concept of introducing disfluency with extreme caution. There are other ways to encourage readers to engage with text that do not make reading more difficult.

You may feel that it was unnecessary to read quite so much about psychological processes involved in reading or the detail of experiments. If so, I expect you have skipped over these parts or skimmed them. The book is designed to enable you to do this, to choose your own reading strategy. If you developed an interest in how we read and how research is carried out, I hope you will pursue this interest in the future. We need to update our research knowledge base to keep track with changes in screen-based technologies and different reading habits. As a typographic or graphic designer, you can make an invaluable contribution, especially if you are open to collaborating with other disciplines.

Bibliography

Arditi, A., & Cho, J. (2007). Letter case and text legibility in normal and low vision. *Vision Research, 47*(19), 2499–2505. https://doi.org/10.1016/j.visres.2007.06.010

Bailey, I. L., & Lovie, J. E. (1976). New design principles for visual-acuity letter charts. *American Journal of Optometry and Physiological Optics, 53*(11), 740–745. https://doi.org/10.1097/00006324-197611000-00006

Bartram, D. (1982). The perception of semantic quality in type: Differences between designers and non-designers. *Information Design Journal, 3*(1), 38–50. https://doi.org/10.1075/idj.3.1.04bar

Baudin, F. (1967). 'Miles A. Tinker, *Bases for effective reading'. Journal of Typographic Research, 1*(2), 204-207. https://journals.uc.edu/index.php/vl/article/view/4998/3862

Becker, D., Heinrich, J., von Sichowsky, R., & Wendt, D. (1970). Reader preferences for typeface and leading. *Journal of Typographic Research, 4*(1), 61–66. https://journals.uc.edu/index.php/vl/article/view/5062

Beier, S. (2012). *Reading letters: Designing for legibility.* Amsterdam: BIS Publishers. https://www.researchgate.net/publication/361332642_Reading_Letters_designing_for_legibility

Beier, S. (2016). Letterform research: An academic orphan. *Visible Language, 50*(2), 64–79. https://journals.uc.edu/index.php/vl/article/view/5923

Beier, S., & Dyson, M. C. (2014). The influence of serifs on 'h' and 'i': Useful knowledge from design-led scientific research. *Visible Language, 47*(3), 74–95. https://journals.uc.edu/index.php/vl/article/view/5875

Beier, S., & Larson, K. (2010). Design improvements for frequently misrecognized letters. *Information Design Journal, 18*(2), 118–137. https://doi.org/10.1075/idj.18.2.03bei

Beier, S., & Larson, K. (2013). How does typeface familiarity affect reading performance and reader preference? *Information Design Journal, 20*(1), 16–31. https://doi.org/10.1075/idj.20.1.02bei

Bentley, M. (1921). Leading and legibility. *Psychological Monographs: General and Applied, 30*(3), 48–61. https://psycnet.apa.org/doi/10.1037/h0093140

Bernard, M., Lida, B., Riley, S., Hackler, T., & Janzen, K. (2002). A comparison of popular online fonts: Which size and type is best? *Usability News, 4*(1). http://usabilitynews.org/a-comparison-of-popular-online-fonts-which-size-and-type-is-best/.

Bernard, M., Mills, M., Peterson, M., & Storrer, K. (2001). A comparison of popular online fonts: Which is best and when? [Electronic Version]. *Usability News, 3*(2). http://usabilitynews.org/a-comparison-of-popular-online-fonts-which-is-best-and-when/.

Besner, D., Coltheart, M., & Davelaar, E. (1984). Basic processes in reading: Computation of abstract letter identities. *Canadian Journal of Psychology, 38*(1), 126–134. https://psycnet.apa.org/doi/10.1037/h0080785

Beymer, D., Russell, D., & Orton, P. (2008). An eye tracking study of how font size and type influence online reading. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction* (Vol. 2, pp. 15–18). Liverpool, UK: British Computer Society. https://dl.acm.org/doi/10.5555/1531826.1531831

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press. https://psycnet.apa.org/record/1994-97967-009

Boyarski, D., Neuwirth, C., Forlizzi, J., & Regli, S. H. (1998). A study of fonts designed for screen display. In C. M. Karat (Ed.), CHI 98: *Human Factors in Computing Systems* (pp. 87–94). Reading, MA: Addison-Wesley, ACM Press. https://doi.org/10.1145/274644.274658

Bringhurst, R. (1992). *The elements of typographic style*. Point Roberts, WA Hartley & Marks.

British Standards Institution (2005). BS 7000:6:2005: *Design management systems: Part 6: Managing inclusive design – guide*. London, England: BSI. https://knowledge.bsigroup.com/products/design-management-systems-managing-inclusive-design-guide/standard

Brumberger, E. R. (2003). The rhetoric of typography: The awareness and impact of typeface appropriateness. *Technical Communication, 50*(2), 224–231. https://www.ingentaconnect.com/content/stc/tc/2003/00000050/00000002/art00007

Cattell, J. M. (1886). The time taken up by cerebral operations.*Mind, 11*(43), 377–392. https://www.jstor.org/stable/2247078

Chaparro, B. S., Shaikh, A. D., & Baker, J. R. (2005). Reading online text with a poor layout: Is performance worse? *Usability News*. https://researchinuserexperience.wordpress.com/2005/02/13/reading-online-text-with-a-poor-layout-is-performance-worse/

Chapman–Cook speed of reading test. 1923–1924: Grades 4–8. Educational Test Bureau. Chapman

Chapman-Cook speed of reading test. 1923, 1924. Grades 4–8. Educational Test Bureau, Chapman, J. C.

Chung, S. T. L. (2002). The effect of letter spacing on reading speed in central and peripheral vision. *Investigative Ophthalmology & Visual Science, 43*(4), 1270–1276. https://iovs.arvojournals.org/article.aspx?articleid=2200181

Chung, S. T. L. (2004). Reading speed benefits from increased vertical word spacing in normal peripheral vision. *Optometry and Vision Science, 81*(7), 525–535. https://journals.lww.com/optvissci/Abstract/2004/07000/Reading_Speed_Benefits_from_Increased_Vertical.14.aspx

de Bruijn, D., de Mul, S., & van Oostendorp, H. (1992). The influence of screen size and text layout on the study of text. *Behaviour & Information Technology, 11*(2), 71–78. https://doi.org/10.1080/01449299208924322

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition, 118*(1), 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics, 35*(10), 1297–1326. https://doi.org/10.1080/00140139208967394

Dillon, A. (2004). *Designing usable electronic text* (2nd ed.). Boca Raton: CRC Press. https://doi.org/10.1201/9780367801151

Dillon, A., Kleinman, L., Bias, R., Choi, G. O., & Turnbull, D. (2004). Reading and searching digital documents: An experimental analysis of the effects of image quality on user performance and perceived effort. *Proceedings of the American Society for Information Science and Technology, 41*(1), 267–273. https://doi.org/10.1002/meet.1450410131

Dobres, J., Chahine, N., Reimer, B., Gould, D., Mehler, B., & Coughlin, J. F. (2016). Utilizing psychophysical techniques to investigate the effects of age, typeface design, size, and display polarity on glance legibility. *Ergonomics, 59*(10), 1377–1391. https://doi.org/10.1080/00140139.2015.1137637

Dowding, G. (1966). *Finer points in the spacing and arrangement of type*. London: Wace.

Doyle, J. R., & Bottomley, P. A. (2004). Font appropriateness and brand choice. *Journal of Business Research, 57*(8), 873–880. https://psycnet.apa.org/doi/10.1016/S0148-2963(02)00487-3

Doyle, J. R., & Bottomley, P. A. (2006). Dressed for the occasion: Font-product congruity in the perception of logotype. *Journal of Consumer Psychology, 16*(2), 112–123. https://doi.org/10.1207/s15327663jcp1602_2

Dyson, M. C. (2004). How physical text layout affects reading from screen. *Behaviour & Information*

Technology, 23(6), 377–393. https://doi.org/10.1080/01449290410001715714

Dyson, M. C. (2005). How do we read text on screen? In H. v. Oostendorp, L. Breure & A. Dillon (Eds.), *Creation, use, and deployment of digital information* (pp. 279–306). Mahwah, NJ: Lawrence Erlbaum Associates. https://www.routledge.com/Creation-Use-and-Deployment-of-Digital-Information/Oostendorp-Breure-Dillon/p/book/9780805845877

Dyson, M. C., & Beier, S. (2016). Investigating typographic differentiation: Italics are more subtle than bold for emphasis. *Information Design Journal, 22*(1), 3–18. https://doi.org/10.1075/idj.22.1.02dys

Dyson, M. C., & Kipping, G. J. (1997). The legibility of screen formats: Are three columns better than one? *Computers & Graphics, 21*(6), 703–712. https://doi.org/10.1016/S0097-8493(97)00048-4

Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: Does harder-to-read mean better-to-understand? *Applied Cognitive Psychology, 28*(4), 488–501. https://doi.org/10.1002/acp.3004

Fiset, D., Blais, C., Arguin, M., Tadros, K., Éthier-Majcher, C., Bub, D., et al. (2009). The spatio-temporal dynamics of visual letter recognition. *Cognitive Neuropsychology, 26*(1), 23–35. https://doi.org/10.1080/02643290802421160

Fiset, D., Blais, C., Éthier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for identification of uppercase and lowercase letters. *Psychological Science, 19*(11), 1161–1168. https://doi.org/10.1111/j.1467-9280.2008.02218.x

Foster, J. J. (1970). A study of the legibility of one- and two-column layouts for bps publications. *Bulletin of the British Psychological Society, 23*, 113–114. https://www.researchgate.net/publication/313142619_A_study_of_the_legibility_of_one_and_two_column_layouts_for_BPS_publications

Foster, J. J. (1980). *Legibility research 1972–1978: A summary*. London: Graphic Information Research Unit, Royal College of Art.

Fox, D., Chaparro, B. S., & Merkle, E. (2007). Examining legibility of the letter “e” and number “0” using classification tree analysis. *Usability News, 9*(2). https://www.researchgate.net/publication/254847470_Examining_Legibility_of_the_Letter_e_and_Number_0_Using_Classification_Tree_Analysis

Frutiger, A. (1998). *Signs and symbols: Their design and meaning*. London: Ebury Press.

Gill, E. (1931). *An essay on typography*. London: Dent.

Grabinger, R. S. (1993). Computer screen designs: Viewer judgements. *Educational Technology and Research Development, 41*(2), 35–73. https://doi.org/10.1007/BF02297311

Grainger, J. (2016). Orthographic processing and reading. *Visible Language, 50*(2), 132–153. https://journals.uc.edu/index.php/vl/article/view/5925

Grainger, J., & Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe? *Trends in Cognitive Sciences, 8*(2), 58–59. https://psycnet.apa.org/doi/10.1016/j.tics.2003.11.006

Grainger, J., Dufau, S., & Ziegler, J. C. (2016). A vision of reading. *Trends in Cognitive Sciences, 20*(3), 171–179. https://psycnet.apa.org/doi/10.1016/j.tics.2015.12.008

Grainger, J., Lété, B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition, 123*(2), 280–292. https://doi.org/10.1016/j.cognition.2012.01.003

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in Cognitive Sciences, 12*(10), 381–387. https://psycnet.apa.org/doi/10.1016/j.tics.2008.06.006

Gregory, M., & Poulton, E. C. (1970). Even versus uneven right-hand margins and the rate of comprehension in reading. *Ergonomics, 13*(4), 427–434. https://doi.org/10.1080/00140137008931157

Haber, R. N., & Schindler, R. M. (1981). Error in proofreading – evidence of syntactic control of letter processing. *Journal of Experimental Psychology: Human Perception and Performance, 7*(3), 573–579. https://psycnet.apa.org/doi/10.1037/0096-1523.7.3.573

Harris, J. (1973). Confusions in letter recognition. *Printing Technology, 17*(2), 29–34.

Hartley, J., & Trueman, M. (1983). The effects of headings in text on recall, search and retrieval. *British Journal of Educational Psychology, 53*(JUN), 205–214. https://psycnet.apa.org/doi/10.1111/j.2044-8279.1983.tb02551.x

Hartley, J., Burnhill, P., & Davis, L. (1978). The effects of line length and paragraph denotation on the retrieval of information from prose text. *Visible Language, 12*(2), 183–194. https://journals.uc.edu/index.php/vl/article/view/5255

Hartley, J., Burnhill, P., & Fraser, S. (1974). Typographical problems of journal design. *Applied Ergonomics, 5*(1), 15–20. https://psycnet.apa.org/doi/10.1016/0003-6870(74)90253-1

Hazlett, R. L., Larson, K., Shaikh, A. D., & Chaparro, B. S. (2013). Two studies on how a typeface congruent with content can enhance onscreen communication. *Information Design Journal, 20*(3), 207–219. https://doi.org/10.1075/idj.20.3.02haz

Huey, E. B. (1908/1968). *The psychology and pedagogy of reading*. Cambridge, Massachusetts: MIT Press. https://psycnet.apa.org/record/1908-10406-000

Hughes, L., & Wilkins, A. (2000). Typography in children's reading schemes may be suboptimal: Evidence from measures of reading rate. *Journal of Research in Reading, 23*(3), 314–324. http://dx.doi.org/10.1111/1467-9817.00126

Hyönä, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. *Learning and Instruction 14*(2), 131–152. https://doi.org/10.1016/j.learninstruc.2004.01.001

Jäkel, F., Singh, M., Wichmann, F. A., & Herzog, M. H. (2016). An overview of quantitative approaches in Gestalt perception. *Vision Research, 126*, 3–8. https://doi.org/10.1016/j.visres.2016.06.004

Joo, S. J., White, A. L., Strodtman, D. J., & Yeatman, J. D. (2018). Optimizing text for an individual's visual system: The contribution of visual crowding to reading difficulties. *Cortex, 103*, 291–301. https://doi.org/10.1016/j.cortex.2018.03.013

Juni, S., & Gross, J. S. (2008). Emotional and persuasive perception of fonts.*Perceptual and Motor Skills, 106*(1), 35–42. https://doi.org/10.2466/pms.106.1.35-42

Keage, H. A. D., Coussens, S., Kohler, M., Thiessen, M., & Churches, O. F. (2014). Investigating letter recognition in the brain by varying typeface: An event-related potential study. *Brain and Cognition, 88*, 83–89. https://doi.org/10.1016/j.bandc.2014.05.001

Kelly, G. (1955). *The psychology of personal constructs*. New York: Norton. https://psycnet.apa.org/record/1956-04524-000

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363–394. https://psycnet.apa.org/doi/10.1037/0033-295X.85.5.363

Koffka, K. (1935). *Principles of Gestalt psychology*. London: Routledge & Kegan Paul. https://psycnet.apa.org/record/1935-03991-000

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright Publishing Corporation. https://psycnet.apa.org/record/1947-03787-000

Kolers, P. A., Duchnicky, R. L., & Ferguson, D. C. (1981). Eye movement measurement of readability of CRT displays. *Human Factors, 23*(5), 517–527. https://doi.org/10.1177/001872088102300502

Köpper, M., Mayr, S., & Buchner, A. (2016). Reading from computer screen versus reading from paper: Does it still make a difference? *Ergonomics, 59*(5), 615–632. https://doi.org/10.1080/00140139.2015.1100757

Lanthier, S. N., Risko, E. R., Stolzh, J. A., & Besner, D. (2009) Not all visual features are created equal: Early processing in letter and word recognition. *Psychonomic Bulletin & Review, 16*(1), 67–73. https://psycnet.apa.org/doi/10.3758/PBR.16.1.67

Larson, K. (2004). The science of word recognition. http://www.microsoft.com/typography/ctfonts/WordRecognition.aspx

Larson, K., & Carter, M. (2016). Sitka: A collaboration between type design and science. In M. C. Dyson & C. Y. Suen (Eds.), Digital fonts and reading (pp. 37–53). New Jersey: World Scientific. http://dx.doi.org/10.1142/9789814759540_0003

Larson, K., Hazlett, R. L., Chaparro, B. S., & Picard, R. W. (2006) Measuring the aesthetics of reading. In People and computers XX: Proceedings of HCI 2006, (Vol. 1, pp. 41–56): British Computer Society. https://link.springer.com/chapter/10.1007/978-1-84628-664-3_4

Legge, G. E. (2007). *Psychophysics of reading in normal and low vision*. Mahwah, New Jersey: Lawrence Erlbaum Associates. https://psycnet.apa.org/record/2006-20917-000

Legge, G. E., & Bigelow, C. A. (2011) Does print size matter for reading? A review of findings from vision science and typography. *Journal of Vision, 11*(5):8, 1–22. https://doi.org/10.1167/11.5.8

Legros, L. A., & Grant, J. C. (1916). *Typographical printing-surfaces: The technology and mechanism of their production*. London: Longmans, Green.

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning, 11*(1), 89–105. https://psycnet.apa.org/doi/10.1007/s11409-015-9149-z

Lewis, C., & Walker, P. (1989). Typographic influences on reading. *British Journal of Psychology, 80*(2), 241–257. https://psycnet.apa.org/doi/10.1111/j.2044-8295.1989.tb02317.x

Licko, Z. (1990). Do you read me? *Emigre, 15*, 8–13. https://www.emigre.com/Essays/ZuzanaLicko/Emigre15

Ling, J., & van Schaik, P. (2006). The influence of font type and line length on visual search and information retrieval in web pages. *International Journal of Human-Computer Studies, 64*(5), 395–404. https://doi.org/10.1016/j.ijhcs.2005.08.015

Ling, J., & van Schaik, P. (2007). The influence of line spacing and text alignment on visual search of web pages. *Displays, 28*, 60–67. https://psycnet.apa.org/doi/10.1016/j.displa.2007.04.003

Lonsdale, M. (2007). Does typographic design of examination materials affect performance? *Information Design Journal, 15*(2), 114–138. https://doi.org/10.1075/idj.15.2.04lon

Lonsdale, M., Dyson, M. C., & Reynolds, L. (2006). Reading in examination-type situations: The effects of text layout on performance. *Journal of Research in Reading, 29*(4), 433–453. https://doi.org/10.1111/j.1467-9817.2006.00317.x

Lund, O. (1999). *Knowledge construction in typography: The case of legibility research and the legibility of sans serif typefaces*. Unpublished PhD thesis, University of Reading, UK. https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.301973

Macaya, M., & Perea, M. (2014). Does bold emphasis facilitate the process of visual-word recognition? *Spanish Journal of Psychology, 17*(e2), 1–5. https://doi.org/10.1017/sjp.2014.2

MacDonald-Ross, M., & Waller, R. (1975). Criticism, alternatives and tests: A conceptual framework for improving typography. *Programmed Learning and Educational Technology, 12*(2), 75–83. https://doi.org/10.1080/1355800750120202

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5*(1), 58–70. https://doi.org/10.1167/5.1.6

McClelland, J. L., & Rumelhart, D. 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. https://doi.org/10.1037/0033-295x.88.5.375

Monk, A. F., & Hulme, C. (1983). Errors in proofreading: Evidence for the use of word shape in word recognition. *Memory & Cognition, 11*(1), 16-23. https://doi.org/10.3758/bf03197657

Moret-Tatay, C., & Perea, M. (2011). Do serifs provide an advantage in the recognition of written words? *Journal of Cognitive Psychology, 23*(5), 619–624. https://doi.org/10.1080/20445911.2011.546781

Morris, R. A., Aquilante, K., Yager, D., & Bigelow, C. (2002). P-13: Serifs slow RSVP reading at very small sizes, but don't matter at larger sizes. *SID Symposium Digest of Technical Papers, 33*(1), 244–247. http://dx.doi.org/10.1889/1.1830242

Moys, J. L. (2014a). Investigating readers' impressions of typographic differentiation using repertory grids. *Visible Language, 47*(3), 96–123. https://journals.uc.edu/index.php/vl/article/view/5876

Moys, J. L. (2014b). Typographic layout and first impressions – testing how changes in text layout influence readers' judgments of documents. *Visible Language, 48*(1), 41–67. https://journals.uc.edu/index.php/vl/article/view/5879

Moys, J. L., Loveland, P., & Dyson, M. C. (2018). Eink versus paper: Exploring the effects of medium and typographic quality on recall and reading speed. *Visible Language, 52*(3), 74–95. https://journals.uc.edu/index.php/vl/article/view/4634

Muter, P., & Maurutto, P. (1991). Reading and skimming from computer screens and books: The paperless office revisited? *Behaviour & Information Technology, 10*(4), 257–266. https://doi.org/10.1080/01449299108924288

Nelson-Denny reading test. (1981). Chicago, Illinois: The Riverside Publishing Company.

New, B., & Grainger, J. (2011). On letter frequency effects. *Acta Psychologica, 138*(2), 322–328. https://doi.org/10.1016/j.actpsy.2011.07.001

Ovink, G. W. (1938). *Legibility, atmosphere-value and forms of printing type*. Leiden: Sijthoff.

Paap, K. R., Newsome, S. L., & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 413–428. https://psycnet.apa.org/doi/10.1037/0096-1523.10.3.413

Paterson, D. G., & Tinker, M. A. (1932). Studies of typographical factors influencing speed of reading: X. Style of type face. *Journal of Applied Psychology, 16*(6), 605–613. https://psycnet.apa.org/doi/10.1037/h0070644

Paterson, D. G., & Tinker, M. A. (1940). *How to make type readable*. New York: Harper and Row.

Pelli, D. G., & Tillman, K. A. (2007). Parts, wholes, and context in reading: A triple dissociation.*Plos One, 2*(8), e680. https://doi.org/10.1371/journal.pone.0000680

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006) Feature detection and letter identification. *Vision Research 46*(28), 4646–4674. https://doi.org/10.1016/j.visres.2006.04.023

Perea, M. (2012). Revisiting Huey: On the importance of the upper part of words during reading. *Psychonomic Bulletin & Review, 19*(6), 1148–1153. https://doi.org/10.3758/s13423-012-0304-0

Perea, M. (2013). Why does the APA recommend the use of serif fonts?*Psicothema, 25*(1), 13–17. https://doi.org/10.7334/psicothema2012.1419

Perea, M., & Gomez, P. (2012a). Increasing interletter spacing facilitates encoding of words. *Psychonomic Bulletin & Review, 19*(2), 332–338. https://psycnet.apa.org/doi/10.3758/s13423-011-0214-6

Perea, M., & Gomez, P. (2012b). Subtle increases in interletter spacing facilitate the encoding of words during normal reading. *Plos One, 7*(10), e47568. https://doi.org/10.1371/journal.pone.0047568

Perea, M., Giner, L., Marcet, A., & Gomez, P. (2016). Does extra interletter spacing help text reading in skilled adult readers? *Spanish Journal of Psychology, 19*, e26. https://psycnet.apa.org/record/2016-25768-001

Perea, M., Moret-Tatay, C., & Gomez, P. (2011). The effects of interletter spacing in visual-word recognition. *Acta Psychologica, 137*(3), 345–351. https://doi.org/10.1016/j.actpsy.2011.04.003

Perea, M., Panadero, V., Moret-Tatay, C., & Gómez, P. (2012). The effects of inter-letter spacing in visual-word recognition: Evidence with young normal readers and developmental dyslexics. *Learning and Instruction, 22*(6), 420–430. https://doi.org/10.1016/j.learninstruc.2012.04.001

Perea, M., Rosa, E., & Marcet, A. (2017). Where is the locus of the lowercase advantage during sentence reading? *Acta Psychologica, 177*, 30–35. https://doi.org/10.1016/j.actpsy.2017.04.007

Petit, J. P., & Grainger, J. (2002). Masked partial priming of letter perception. *Visual Cognition, 9*(3), 337–353. https://doi.org/10.1080/13506280042000207

Pollatsek, A., Raney, G. E., Lagasse, L., & Rayner, K. (1993) The use of information below fixation in reading and in visual-search. *Canadian Journal of Experimental Psychology–Revue Canadienne de Psychologie Experimentale, 47*(2), 179–200. https://psycnet.apa.org/doi/10.1037/h0078824

Poulton, E. C. (1959). Effects of printing types and formats on the comprehension of scientific journals. *Nature, 184*, 1824–1825. https://doi.org/10.1038/1841824a0

Poulton, E. C. (1965). Letter differentiation and rate of comprehension in reading. *Journal of Applied Psychology, 49*(5), 358–362. https://doi.org/10.1037/h0022461

Poulton, E. C. (1967). Searching for newspapers headlines printed in capitals or lower-case letters. *Journal of Applied Psychology, 51*(5), 417–425. https://psycnet.apa.org/doi/10.1037/h0025098

Poynor, R. (1999). The great type debate rages on. *Graphis, 319*, 14,16. https://www.graphis.com/store_/product/issue-319/

Pyke, R. L. (1926). *Report on the legibility of print*. Her Majesty's Stationery Office.

Rayner, K. (1986). Eye-movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41*(2), 211–236. https://doi.org/10.1016/0022-0965(86)90037-8

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Hillsdale, New Jersey: Lawrence Erlbaum. https://psycnet.apa.org/record/1989-97284-000

Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding wrods with jubmled lettres – there is a cost. *Psychological Science, 17*(3), 192–193. https://doi.org/10.1111/j.1467-9280.2006.01684.x

Rehe, R. F. (1979). *Typography: How to make it most legible*. Carmel, Indiana: Design Research Publications.

Reimer, B., Mehler, B., Dobres, J., Coughlin, J. F., Matteson, S., Gould, D., et al. (2014). Assessing the impact of typeface design in a text-rich automotive user interface. *Ergonomics, 57*(11), 1643–1658. https://doi.org/10.1080/00140139.2014.940000

Rey, A., Dufau, S., Massol, S., & Grainger, J. (2009). Testing computational models of letter perception with item-level event-related potentials. *Cognitive Neuropsychology, 26*(1), 7–22. https://doi.org/10.1080/09541440802176300

Reynolds, L. (1984). The legibility of printed scientific and technical information. In R. Easterby & H. Zwaga (Eds.), *Information design: The design and evaluation of signs and printed material* (pp. 187–208). Chichester: John Wiley.

Reynolds, L., & Simmonds, D. (1984). The legibility of type. In *Presentation of data in science* (pp. 1–19): Springer Netherlands. https://doi.org/10.1007/978-94-009-8354-0_1

Roethlein, B. E. C. (1912). The relative legibility of different faces of printing types. *The American Journal of Psychology, 23*(1), 1–36. https://doi.org/10.2307/1413112

Rosa, E., Perea, M., & Enneson, P. (2016). The role of letter features in visual-word recognition: Evidence from a delayed segment technique. *Acta Psychologica, 169*, 133–142. https://doi.org/10.1016/j.actpsy.2016.05.016

Rudgard-Redsell, P. (2014). *The perceived versus actual usability of mobile navigation designs*. Unpublished BA dissertation, University of Reading, UK.

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review, 81*(2), 99–118. https://psycnet.apa.org/doi/10.1037/h0036117

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacognition and Learning, 11*(1), 57–70. https://doi.org/10.1007/s11409-015-9151-5

Sanford, E. C. C. (1888). The relative legibility of the small letters. *The American Journal of Psychology, 1*(3), 402–435. https://doi.org/10.2307/1411012

Sanocki, T. (1987). Visual knowledge underlying letter perception: Font-specific schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance, 13*(2), 267–278. https://psycnet.apa.org/doi/10.1037/0096-1523.13.2.267

Sanocki, T. (1988). Font regularity constraints on the process of letter recognition *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 472–480. https://doi.org/10.1037//0096-1523.14.3.472

Shaikh, A. D., & Chaparro, B. (2016). Perception of fonts: Perceived personality traits and appropriate uses. In M. C. Dyson & C. Y. Suen (Eds.), *Digital fonts and reading* (pp. 226–247). New Jersey: World Scientific. https://doi.org/10.1142/9789814759540_0013

Shaikh, A. D., & Fox, D. (2008). Does the typeface of a resume impact our perception of the applicant. https://www.semanticscholar.org/paper/Does-the-Typeface-of-a-Resume-Impact-Our-Perception-Shaikh-Fox/b2b96fd092271f1aeb4001f5db09f2b746f6adfa

Sheedy, J. E., Subbaram, M., Zimmerman, A., & Hayes, J. R. (2005). Text legibility and the letter superiority effect. *Human Factors, 47*(4), 797–815. https://doi.org/10.1518/001872005775570998

Sheedy, J., Tai, Y.-C., Subbaram, M., Gowrisankaran, S., & Hayes, J. (2008). ClearType sub-pixel text rendering: Preference, legibility and reading performance. *Displays, 29*(2), 138–151.

https://doi.org/10.1016/j.displa.2007.09.016

Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority – does the medium, screen versus paper, affect performance even with brief tasks? *Metacognition and Learning, 11*(1), 15–33. https://doi.org/10.1007/s11409-015-9150-6

Siegenthaler, E., Wurtz, P., Bergamin, P., & Groner, R. (2011). Comparing reading processes on e-ink displays and print. *Displays, 32*(5), 268–273. https://doi.org/10.1016/j.displa.2011.05.005

Slattery, T. J., & Rayner, K. (2010). The influence of text legibility on eye movements during reading. *Applied Cognitive Psychology, 24*(8), 1129–1148. https://doi.org/10.1002/acp.1623

Slattery, T. J., & Rayner, K. (2013). Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text. *Attention Perception & Psychophysics, 75*, 1275–1292.https://doi.org/10.3758/s13414-013-0463-8

Slattery, T. J., Yates, M., & Angele, B. (2016). Interword and interletter spacing effects during reading revisited: Interactions with word and font characteristics. *Journal of Experimental Psychology: Applied, 22*(4), 406–422. https://psycnet.apa.org/doi/10.1037/xap0000104

Sloan, L. (1959). New test charts for the measurement of visual acuity at far and near distances. American *Journal of Opthamology, 48*(6), 807–813. https://doi.org/10.1016/0002-9394(59)90626-9

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science, 19*(10), 986–988. https://doi.org/10.1111/j.1467-9280.2008.02189.x

Song, H., & Schwarz, N. (2010). If it's easy to read, it's easy to do, pretty, good, and true.*The Psychologist, 23*(2), 108–111. https://www.bps.org.uk/psychologist/if-its-easy-read-its-easy-do-pretty-good-and-true

Spencer, H. (1968). *The visible word*. London: Royal College of Art.

Spencer, H. (1970). The future role of the printed word. In*Typographic opportunities in the computer age: Papers of the 11th congress of the Association Typographique Internationale Prague*, June 1969 (pp. 71–75). Prague: Typografia Prague.

Stroop, J. R. (1935/1992). Studies of interference in serial verbal reactions (reprinted from Journal of Experimental Psychology, vol 18, pp. 643–662, 1935). *Journal of Experimental Psychology: General, 121*(1), 15–23. https://psycnet.apa.org/doi/10.1037/0096-3445.121.1.15

Strukelj, A., Scheiter, K., Nystrom, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning, 11*(1), 71–88. https://doi.org/10.1007/s11409-015-9146-2

Tannenbaum, P. H., Jacobson, H. K., & Norris, E. L. (1964). An experimental investigation of typeface connotations. *Journalism Quarterly, 41*(1), 65–73. https://doi.org/10.1177/107769906404100108

Tejero, P., Perea, M., & Jimenéz, M. (2014). Is there a genuine advantage to the upper part of words during lexical access? Evidence from the Stroop task. *Memory & Cognition, 42*(5), 834–841. https://doi.org/10.3758/s13421-013-0390-x

Thiessen, M., Kohler, M., Churches, O., Coussens, S., & Keage, H. (2015). Brainy type: A look at how the brain processes typographic information. *Visible Language, 49*(1–2), 175–188. https://journals.uc.edu/index.php/vl/article/view/5903

Thompson, J. (2009). *The relationship between perceived and actual legibility*. Unpublished BA dissertation, University of Reading, UK.

Tinker, M. A. (1928). The relative legibility of the letters, the digits, and of certain mathematical signs. Journal of General Psychology, 1, 472–496. https://doi.org/10.1080/00221309.1928.9918022

Tinker, M. A. (1947). *The Tinker speed of reading test*. Minneapolis, MN: The University of Minnesota Press.

Tinker, M. A. (1963). *Legibility of print*. Ames, Iowa: Iowa State University Press.

Tinker, M. A. (1965). *Bases for effective reading*. Minneapolis: Lund Press.

Tracy, W. (1982). Other replies to Donald E. Knuth's article. "The concept of a meta-font". *Visible Language, 16*(4), 353–355. https://journals.uc.edu/index.php/vl/issue/view/360

Tracy, W. (1986). *Letters of credit: A view of type design*. London: Gordon Fraser.

Twyman, M. L. (1981). Typography without words. *Visible Language, 15*(1), 5–12. https://journals.uc.edu/index.php/vl/article/view/5310

Wade, N. J., Tatler, B. W., & Heller, D. (2003). Dodge-ing the issue: Dodge, Javal, Hering, and the measurement of saccades in eye-movement research. Perception, 32(7), 793–804. https://doi.org/10.1068/p3470

Walker, S., & Reynolds, L. (2003). Serifs, sans serifs and infant characters in children's reading books. *Information Design Journal, 11*(3), 106–122. http://dx.doi.org/10.1075/idj.11.2.04wal

Warde, B. (1930/2009). The crystal goblet or why printing should be invisible. In H. Armstrong (Ed.), *Graphic design theory: readings from the field* (pp. 39–43). New York, US: Princeton Architectural Press.

Wendt, D. (1970). By what criteria are we to judge legibility? In *Typographic opportunities in the computer age: Papers of the 11th congress of the Association Typographique Internationale Prague*, June 1969 (pp. 42–46). Prague: Typografia Prague.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II.*Psychologische Forschung, 4*, 301–350. https://doi.org/10.1007/BF00410640

Wheildon, C. (1986). *Communicating or just making pretty shapes*. Sydney, Australia: Newspaper Advertising Bureau of Australia Limited.

Wheildon, C. (1995). *Type & layout*. Berkeley, California: Strathmoor Press.

Williams, T. R., & Spyridakis, J. H. (1992). Visual discriminability of headings in text. IEEE Transactions on Professional Communication, 35(2), 64–70. https://doi.org/10.1002/9781119134633.ch18

Williamson, H. (1983). *Methods of book design: The practice of an industrial craft* (3rd ed.). London: Yale University Press.

Wrolstad, M. E. (1970). Methods of research into legibility and intelligibility. In *Typographic opportunities in the computer age: Papers of the 11th congress of the Association Typographique Internationale Prague*, June 1969 (pp. 36–41). Prague: Typografia Prague.

Youngman, M., & Scharff, L. (1998). Text width and margin width influences. Paper presented at the *Southwestern Psychological Association Conference 1998*. http://www.lieb.com/Readings/Width.pdf

Yu, D., Cheung, S.-H., Legge, G. E., & Chung, S. T. L. (2007) Effect of letter spacing on visual span and reading speed. *Journal of Vision, 7*(2), 1–10. https://doi.org/10.1167/7.2.2

Zachrisson, B. (1965). *Studies in the legibility of printed text*. Stockholm: Almqvist and Wiksell.

Zachrisson, B. (1970). The problem of congeniality in typography. In *Typographic opportunities in the computer age: Papers of the 11th congress of the Association Typographique Internationale Prague, June 1969* (pp. 47–51). Prague: Typografia Prague.

Acknowledgements

My thanks go to:

- Sofie Beier for providing me with her typefaces for some of my illustrations
- Jeanne-Louise Moys for sharing illustrations of her test material and reading a draft
- David Brezina for answering my queries on typefaces, terminology and more
- Carlos Pérez Cerón who did an amazing job in interpreting my suggestions
- María González de Cossío for sensing that this project was just what I needed and who was patient and supportive throughout
- Kevin Larson for his detailed read and helpful feedback on the first edition