Applying psychological theory to typography: is how we perceive letterforms special?

Book or Report Section

Accepted Version

Applying psychological theory to typography: is how we perceive letterforms special?

1 Summary

My research looks for parallels between our perception of letters and typefaces (by readers and designers) and other perceptual activities. Because I am concerned with perception, I have drawn upon examples of research into other areas of perception, both visual and auditory, i.e. the perception of faces, music, and speech. These research studies suggest to me avenues to explore in relation to how we perceive visual forms, and also stimulate ideas concerning particular methods of investigation. This approach leads to novel experiments within the field of typography.

As a teacher, I am interested in how we train students in the visual discriminations that are required of typographers and what characterizes typographic expertise. But I am also concerned with the more general question of how we, as readers, recognize letters regardless of the typeface which changes their visual form. Although readers are not typographic experts, they are experts in letter perception, just as listeners are expert in speech perception. How do we process letters and how might typographers' perceptual skills differ from those of readers?

The outcomes of this research are intended to have theoretical implications, contributing to models of reading, but also to have practical applications: a better understanding of reading could inform the teaching of reading and reading materials; gaining insight into how we process typefaces could feed into their design; identifying the nature of typographers' perceptual skills could facilitate training. Overall, my aim is to understand how we perceive letterforms.

2. Introduction

2.1 Why this question?

One debate among psychologists is whether we have a few general mechanisms that we apply to our various activities or more specialised mechanisms that have adapted to our particular needs. General mechanisms provide a simpler, more parsimonious explanation of how we perceive the world and can be intuitively appealing. For example, my PhD thesis looked for an explanation as to why we notice the peaks and troughs in a melody (the high and low notes) through a visual analogy. Isn't it likely that the extremes of a visual display also stand out and we might therefore apply a similar process when listening and looking?[1] It would seem to be more efficient for us to re-use perceptual mechanisms in different circumstances, rather than switch depending on the particular activity.

But the alternative argument is that there may be activities that require more specialization and particular skills may be acquired, such as learning to recognize different faces or understanding speech. Within these areas of research typical questions are "Is face processing special?" or "Is speech special?" So have we developed perceptual mechanisms that are different from those we use for processing objects in general or listening to other sounds?

In my work, I am considering these two explanations of how we perceive variation in typography. Is the way in which readers process typefaces special, or are there parallels in other forms of perception? In addition, I am exploring how typographic expertise may be distinct from letter expertise, i.e. how to characterise perceptual abilities resulting from typographic design training.

Quite a few of my studies therefore recruit students from the Department in which I teach[2] in order to explore the effects of training in typographic design. For example, do typographers perceive typefaces in a different manner to non-typographers? If they do, this may have implications for how we may more effectively and efficiently teach typography and encourage the development of particular ways of analyzing and discriminating visual characteristics of text. Typographers will also need to be aware that their perceptions may not match those of the users of their designs and they must therefore take account of this, e.g. through user testing.

My research was initially motivated by encountering models of reading and experiments conducted to test such models that appeared to avoid explanations of how variations in fonts[3] that are typical of our normal reading material are handled. For example, McClelland and Rumelhart's Interactive Activation Model (1981: 383) describes detectors for visual features of letters, and assumes that "the basic results do not depend on the particular font used" but also acknowledges that this "skirts several fundamental issues about the lower levels of processing".

These observations on the lack of attention to fonts may be regarded as the "natural inclination of a person to think the details of his craft matter" (Oden 1989: 145). This may be true, but a more compelling argument for pursuing my research has been to further understand one component of the complex skill of reading: the early stages of reading, where the lower level of processing occurs. To recognise a word, a reader must have processed the letters to create abstract letter identities (Besner, Coltheart, and Davelaar 1984; Grainger, Rey and Dufau 2008). A skilled reader recognises most words within a fraction of a second despite the letters being in different fonts (Rayner and Pollatsek 1989: 3). With relative ease, we translate variant visual forms (such as different fonts and cases) into invariant representations (Figure 1).
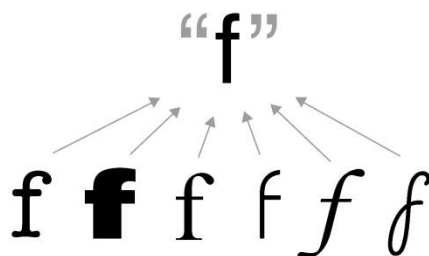

Figure 1:
Letters in different fonts are converted into a single letter identity

2.2   What informs my approach?

The means by which we recognise letters in different fonts has been researched and is described as "font tuning". This process whereby the particular characteristics of a font can be used to facilitate letter recognition was proposed and extensively explored by Sanocki (1987, 1988a, 1988b, 1991a, 1991b, 1992). Fonts generally have a consistency in their design and Sanocki suggests that the perceptual system can become tuned to a particular font over time and develop a set of font parameters. These parameters are used to transform font-specific input into font-invariant form. Sanocki conducted a series of experiments to address how variations in visual forms are translated into font invariant forms (Figure 1) and produced the reliable finding that mixing fonts leads to less efficient letter identification than using the same fonts. These results have been variously described by Sanocki as a "font-regularity effect" (1987) and "font-consistency effect" (1991b).

A similar skill to font tuning is perceiving and understanding speech from many different talkers with considerable variation in the acoustic properties of speech (Martin et al. 1989: 676). The acoustic pattern of a vowel from one talker overlaps with the acoustic pattern of a different vowel spoken by someone else. Yet we are able to identify vowels correctly across different talkers. Although commonly described in terms of the opposite effect (i.e. variability), a similar regularity is found in vowel identification when mixed talkers are compared with a single talker (e.g. Assmann, Nearey and Hogan 1982; Mullennix, Pisoni and Martin 1989). If the talker is fixed, vowel identification is improved.

In looking at how we process letters, I am asking whether there are specialised skills or the application of more general visual and auditory perceptual mechanisms. My research objectives are therefore theoretically oriented, but could have a range of practical applications in for example, informing the design of material for beginner readers, more generally improving ease of reading, or training typographers (as described above).

2.3   How I research the question

2.3.1 Rationale

The methods that I have applied to looking at how we process letters and typefaces draw upon psychological techniques that are designed to generate reliable and objective measures of the processes, and that remain broadly within the paradigms used for related research (e.g. face perception). This approach to revealing our perceptions is necessary because we are not able to accurately report what may take place in less than a second, or we may not be consciously aware of the strategies we use in judging visual material. Adopting a scientific approach to the study of typography complements the craft knowledge of designers and more applied research. Over time, I have developed a desire to bridge the gap between disciplines (e.g. Dyson, 1999a, 1999b, 2000) aided by a better understanding of both "sides". I have done this from within typography inspired by the work of typographically aware psychologists (Prof Pat Wright[4] and Prof James Hartley[5]).

2.3.2 Materials

I conduct experiments that are computer-based so that I can control and record all aspects of the procedure. All material is displayed on screen; the content does not form meaningful text as I use either single paragraphs of dummy text material, or single strings of letters. To create paragraphs of dummy text I use a freely available dynamic text tool which generates random sequences of words.[6] Some of the experiments require judgements of fonts; others require letters to be recognised. Where fonts are being judged, words may be recognisable (see Figure 9a), although some may appear unusual as low frequency words and proper nouns are included. When I am looking at letter processing, this needs to be separated from other aspects of the reading process (e.g. word recognition or sentence comprehension). Using random letter strings removes the semantic (top-down) influences.

2.3.2 Tasks

Of the four studies described in this chapter, two use an identification task which consists of saying whether a paragraph of dummy text or a string of letters (Hamdurefonsiv)[7] is most like one font or another font, i.e. a form of categorisation. The participant is provided with examples of the two fonts (to show what they look like) and must then repeatedly select which of these two is closest to each of the examples they are shown. I use this task for two purposes: to identify the

boundary of the font categories, i.e. where there is a switch from one font to another; to explore whether this boundary changes with prior exposure to one or other of the fonts. These results provide some insight into how typeface characteristics are perceived and the effects of context on this perception.

Three of the four studies include a discrimination task where two paragraphs or two letter strings are judged to be either the same or different. The studies involve two types of judgements: whether the two sets of material are presented in the same or a different font; whether two letter strings contain the same letters or different letters (where one letter has been changed). Figure 3a illustrates an example of different letter strings and Figure 3b, same letter strings. I use this task to further explore how typeface characteristics are perceived, and whether typographers' judgements differ in a qualitative way from those of non-typographers. This task is also designed to reveal whether we make use of typeface characteristics when we identify letters.

In some of my studies described below, the length of time that the material is presented on screen (the viewing duration) is critical. This is because the task must not be too easy or too difficult to be able to demonstrate the effects I am interested in. The viewing duration is therefore set in two ways: (i) through pilot studies or (ii) an adjustment stage prior to the main experiment. In the latter case, I am aiming to tailor the experiment to the particular participant and set a level that is reasonably comfortable for them.[8] In addition to varying the viewing duration, I also can vary the degree of difference between the two sets of material; a larger difference should make the task easier. Participants receive feedback on the accuracy of their responses after each judgement in the form of a tick or cross which hopefully helps to familiarise them with the nature of the task.

2.3.3 Measures of performance

With the identification task, responses are summarised by calculating the percentage identified as one particular font.[9] The discrimination task uses a rating scale from "sure same", "same" to "different", and "sure different". I encourage participants to use all four responses and to distinguish between responses where they are more confident and those where they are less certain. The analysis of these responses is based on Signal Detection Theory (Green and Swets 1966) which when applied to psychology concerns making perceptual decisions based on equivocal evidence (McNicol 1972: 10). When we are asked to make a decision as to whether two strings of letters, presented in quick succession, are the same or different letters, we are unlikely to be certain of our answer. Signal Detection Theory can model this uncertainty by looking at a range of responses, providing us with a more valid measure of our perceptual processes.

The measure I use is p(A)[10] which is an index of discrimination ranging from 0.5 (guessing) to 1 (perfect discrimination); this is a gauge of sensitivity. This measure is separate from another aspect of decision-making which is bias, where one response may be favoured over another. For example, if a participant finds it easier to say that two sets of material are different than to say they are the same[11] they may be biased towards different responses and give more "sure different" responses than "sure same" responses. Their criterion for deciding to respond same may be stricter than for deciding on different. Consequently, they may give the response "different" to some cases where the two sets are the same. If we were to calculate the percent correct responses (which is typically done in psychology experiments), we would naturally count the "different" response to same material as incorrect. But if this participant responds "different" to same material and "sure different" to different material, they are discriminating between same and different but their bias (strict criterion) is interfering with their judgments. Calculating p(A) avoids this interference and

therefore provides a better measure of discrimination because it removes any differences in criterion among participants and within a single participant over time.

Reaction time is commonly measured by psychologists as an indication of whether one condition is easier than another. In my font tuning experiment described below, I have included the time taken to respond same or different because the majority of evidence for the effect I am looking for is from reaction time data. However, this measure can be affected by bias and this cannot be removed. Using the same example as above, a participant may respond more quickly when making a different response because they need accumulate less evidence. Similarly, a participant may decide to take more care when needing to compare two strings of letters in different fonts than two strings in the same font so their responses are slower. For this reason, I use reaction time data only as a secondary measure.

2.3.4 Participants

When recruiting typographers for my studies, I ask for volunteers from final year undergraduates on our Graphic Communication programme, Masters students in Typeface Design, Book Design and Information Design, and PhD students and reimburse them for their time. The studies are submitted for ethical review according to the procedures specified by the University of Reading Research Ethics Committee and are given permission to proceed. In line with the ethical demands of the studies, it is made clear to participants that the study, their participation and the results are not part of any assessment on their programmes.

Non-typographers are recruited from within the University of Reading and are also typically students. Whilst students may not be representative of the general population in many respects, I believe that the processes they employ in perceiving letters are likely to reflect those of most adult readers (of the Latin alphabet).

In one study described below, I compare typographers and non-typographers with the aim of distinguishing between processes used in reading and those used to analyse typefaces.

2.3.5 Critique of my methods

My approach, which uses artificial test conditions, is not always well received by typographers. Although not necessarily made explicit, criteria that are commonly used by typographers to evaluate empirical research within the field are: (i) the extent to which the experimental materials and methods approximate to normal reading practices; (ii) how closely the results fit with what designers know about visual forms. The former view is likely to be shared by applied psychologists but my work has a more theoretical orientation. This requires a means of isolating the component s of the reading process, as described above, to ensure the research has internal validity, even though it may not have ecological validity. Questioning the reliability of the results if they appear surprising is a positive reaction that I encourage. A good scientist questions findings that do not fit with previous knowledge, and investigates further, trying to replicate what they have found. If my results run counter to designers' tacit knowledge, I believe further checking is required.

3.   My studies

3.1   Font tuning

This work tries to directly address how typefaces fit into models of the early stages of reading, where the lower level of processing occurs (as described above).

3.1.1 Updating font tuning studies

The original research into font tuning was started about twenty five years ago, but there has been a minor revival of work (Gauthier et al. 2006; Wong and Gauthier 2005, 2007; Walker 2008). This later work has been able to use more sophisticated fonts than the basic shapes[12] used by Sanocki, due to the proliferation of digital fonts now available to all PC users. A recent review by Sanocki and Dyson (2012) also explores the evidence for font tuning and how this meshes with type design principles.

My research into font tuning takes a similar line to the above authors in looking at whether we tune to current digital fonts, which are obviously of a higher visual quality than the fonts created by Sanocki (Figure 2a), but are also more similar to each other in basic shapes (Figure 2b). I use my discrimination task, which differs from the methods used by all the previous researchers.

Figure 2a: fonts used by Sanocki   Figure 2b: fonts used by Dyson

The participant is asked to say whether a string of letters contains the same letters as the previous string or different letters. To ensure that my participants attempt to identify all the letters, and not just recall the word or letter shapes, the order of the letters are jumbled from one string to the next. If the string contains different letters, just one letter has been changed. A visual mask follows each letter string. This is used to halt the processing of the test material (by having to process the mask) and therefore provides an accurate measure of processing time. Two typical sequences of screens are shown in Figure 3a and 3b.
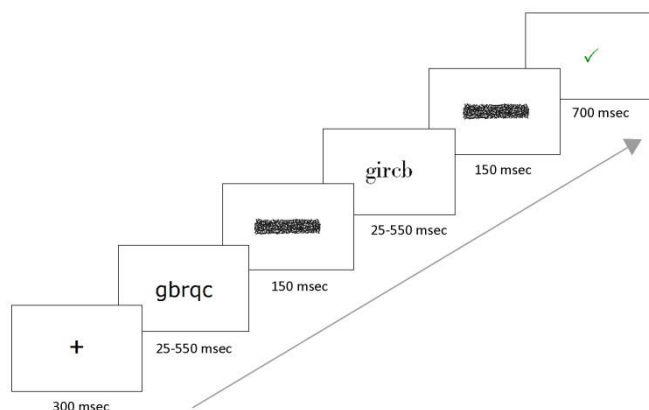

Figure 3a: Sequence of screens in font tuning experiment showing 5-letter strings where the second string is different from the first and displayed in a different font
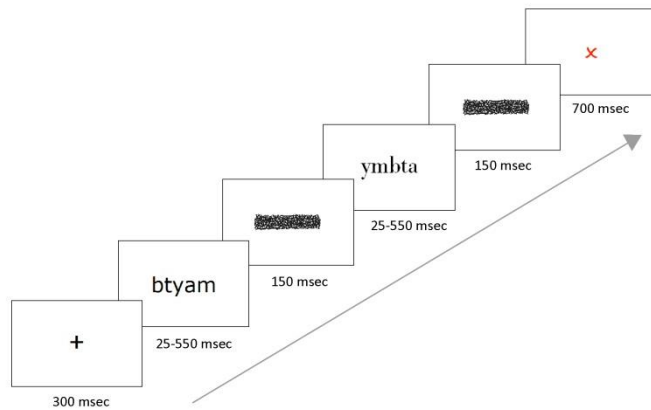
Figure 3b: Sequence of screens in font tuning experiment showing 5-letter strings where the second string is the same as the first and again in a different font

The comparison of interest is whether participants find it easier to identify letters (and thereby discriminate between same and different strings) if the two strings are in the same font as opposed to different fonts. As this effect seemed to be quite elusive, I tried many iterations of the above task using different pairs of fonts, different ways of introducing variation from trial to trial, and different numbers of letters in the string.

3.1.2 Bodoni and Verdana

The results shown in Figure 4 (accuracy) and Figure 5 (reaction time) come from an experiment with non-typographers which uses Bodoni and Verdana, the two fonts shown in Figure 2b, and varies the number of letters in the strings between 4, 5 and 6.
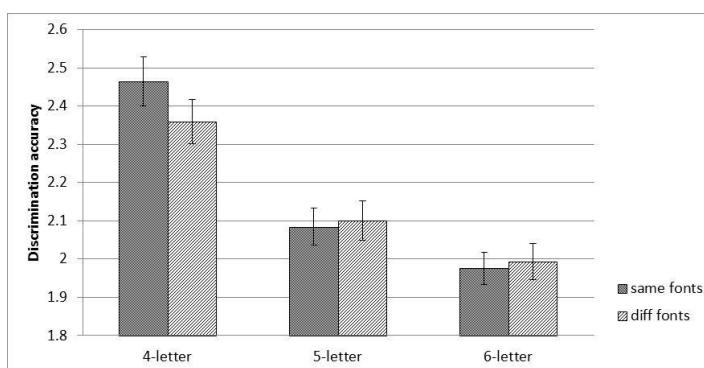

Figure 4: Accuracy of discriminating same and different letter strings when presented in the same or different fonts and varying in length from 4 to 6 letters. The discrimination index p(A) has been transformed for statistical analysis and a perfect score is π ($\approx$3.14159) and chance is $\pi/2$ ($\approx$1.571). The vertical bars show the standard errors of the means which provide an indication of the amount of variation in the scores across participants.

The task is understandably easier, i.e. there is greater accuracy, when there are only 4 letters in the string, but these results also indicate that the advantage for same fonts (the font tuning effect) is only found when there are 4 letters. There would therefore seem to be some limit on the number of letters that are identified more easily by being in the same font (or are more difficult to identify through being in a different font).

7

If we include the time that participants take to respond (Figure 5), the results confirm a font tuning effect, as responses are faster when in the same font. Although the difference seems to be largest when there are only 4 letters, this is not supported statistically.
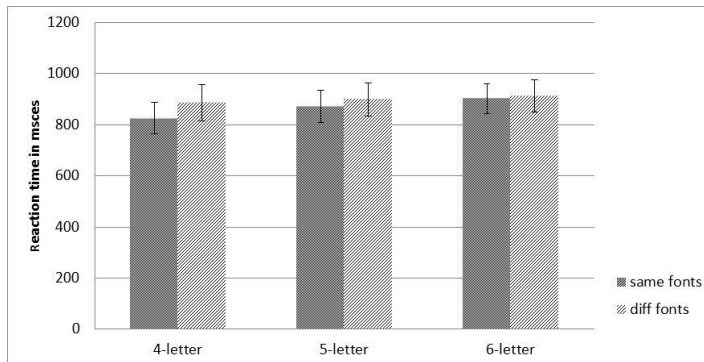

Figure 5: Speed of responding to the letter strings

Processing two strings of letters in the same font therefore facilitates a quicker response, but the consistency in visual form only helps in identifying 4 letters. As introduced in the method section above, faster reaction time might reflect a change in the criterion for responding, e.g. when fonts are the same, there is greater confidence in initiating a response. The accuracy results are not affected by this bias and therefore should provide a more reliable measure of how letters are identified. However, the additional processing required for 5 or 6 letters significantly reduces accuracy. Although performance is still better than guessing, the difficulty of the task may have prevented the consistency in fonts from helping letter identification.[13]

3.1.3 Evidence for font tuning

From these results, I can infer that font tuning occurs with two text fonts in current use, but the effect is particularly sensitive to aspects of the task such as the level of difficulty. The method of adjustment I used attempted to control this by varying the duration of the letter strings, but a compromise was necessary in finding one duration for 4, 5 and 6 letters. Given the greater difficulty with more letters, this compromise may not have been suitable.

I have subsequently demonstrated font tuning in typographers using Garamond and Bodoni in a more complicated experiment with 4-letter strings which involved judgements of letters (as in this study) and of fonts. Sanocki's theory may therefore still be relevant to current fonts, and our perceptual system can become tuned to a particular font, but there may be particular circumstances that elicit this response. Walker (2008) proposes that readers may be able to exercise strategic control over the use of font parameters based on the likelihood of the font changing.

3.2   Categorical perception

Font tuning aims to explain how readers process typefaces. The next study explores typographic expertise.

3.2.1 Cross-modality and within-modality comparisons

A special status for speech perception was originally attributed to the occurrence of categorical perception with speech sounds, but not non-speech (Rosen and Howell 1987: 118). Categorical perception is a psychophysical phenomenon whereby we perceive categories where none physically exist. Although the initial work identified phoneme boundaries from an acoustic continuum (Liberman et al. 1957), categorical perception has now been demonstrated in many visual domains.

In pursuit of some general principles of perception, Bornstein (1987) lists evidence for parallels between visual and auditory categorical perception. A classic example of categorical perception in the visual domain is colour perception where we perceive discrete hues rather than gradually changing ones (Bornstein 1987). Figure 6 illustrates the colour spectrum which is a continuum of light frequencies, which we categorise into blue, green, red, etc.
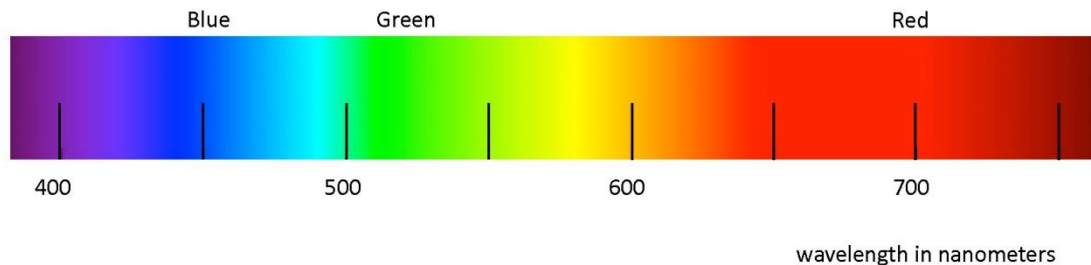
Colour Spectrum

Figure 6: Colour spectrum representing a continuum of light frequencies, categorised as discrete hues

Colour perception corresponds to changes in a single dimension (the wavelength of light) but categorical perception also occurs when there are multidimensional changes. An example of this is categorical perception of faces (Beale and Keil 1995; Campanella et al. 2003). Faces vary in the size and shape of features (eyes, nose, mouth) and also the relative distance between them. Typefaces also vary along multiple dimensions[14]: weight, contrast, proportions, basic shapes, terminals and serifs (Baines and Haslam 2005). Figure 7 illustrates these differences.
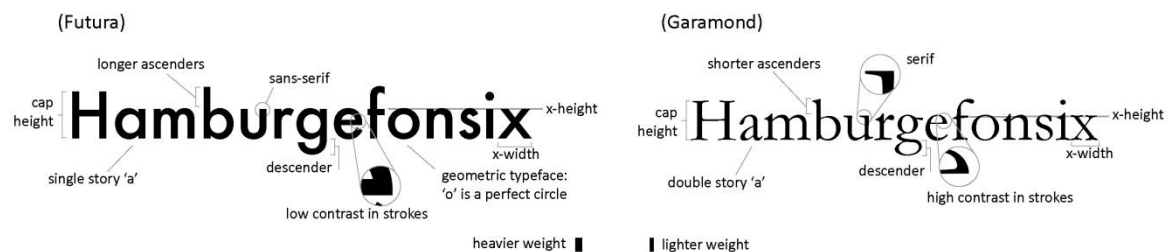

Figure 7: Some of the many ways in which a sans-serif typeface can differ from a serif typeface

The ability to distinguish among typefaces and recognise categories of typefaces (e.g. Humanist, Garalde, Transitional, Didone)[15] or identify individual typefaces is part of typographic training. The influence of training on categorical perception has been demonstrated in the field of music. Trained musicians show categorical perception of musical intervals and those without musical training do not (Burns and Ward 1978). I therefore chose to investigate whether students with some education in typographic or graphic design might perceive typefaces categorically (Dyson 2011).

3.2.2 Creating test material

Because faces and typefaces do not have a natural continuum, i.e. they do not exist as continuously varying shapes in the real world, creating appropriate material for testing is challenging. The interpolation procedure needs to produce a linear continuum with equal steps between points on the continuum, which is not straightforward with multidimensional stimuli (Newell and Bülthoff 2002). A 12-step continuum from Times New Roman to Helvetica (Timevetica) is illustrated in Figure 8a and from Garamond to Bodoni (Garadoni) in Figure 8b. Timevetica was

chosen as the two ends are common examples of a serif and sans-serif typeface; Garadoni uses two serif typefaces which avoids some of the difficulties in the transition from serif to no serif. These continua were developed and modified by trained type designers as the results of the interpolation must be a set of plausible fonts. Equal steps are necessary as a hallmark of categorical perception is that physically equal steps are not perceived as equal. Equal physical differences are perceived as larger or smaller depending on whether or not they are put into the same category.

1 hamburefonsivHBUREFOSI.,
2 hamburefonsivHBUREFOSI.,
3 hamburefonsivHBUREFOSI.,
4 hamburefonsivHBUREFOSI.,
5 hamburefonsivHBUREFOSI.,
6 hamburefonsivHBUREFOSI.,
7 hamburefonsivHBUREFOSI.,
8 hamburefonsivHBUREFOSI.,
9 hamburefonsivHBUREFOSI.,
10 hamburefonsivHBUREFOSI.,
11 hamburefonsivHBUREFOSI.,
12 hamburefonsivHBUREFOSI.,

Figure 8a: Continuum from Times (1) to Helvetica (12)

1 hamdurefonsivHAMDUREFONSIV.,
2 hamdurefonsivHAMDUREFONSIV.,
3 hamdurefonsivHAMDUREFONSIV.,
4 hamdurefonsivHAMDUREFONSIV.,
5 hamdurefonsivHAMDUREFONSIV.,
6 hamdurefonsivHAMDUREFONSIV.,
7 hamdurefonsivHAMDUREFONSIV.,
8 hamdurefonsivHAMDUREFONSIV.,
9 hamdurefonsivHAMDUREFONSIV.,
10 hamdurefonsivHAMDUREFONSIV.,
11 hamdurefonsivHAMDUREFONSIV.,
12 hamdurefonsivHAMDUREFONSIV.,

Figure 8b: Continuum from Garamond (1) to Bodoni (12)

3.2.3 Testing for categorical perception

Both identification and discrimination tasks are used to test for categorical perception. Participants in my experiments are asked to identify examples from along the continuum, saying whether a paragraph is most like Times or Helvetica (Figure 9a) or most like Garamond or Bodoni (Figure 9b). A category boundary is determined where half the responses are from one end of the

continuum and half from the other (Figure 10). For Timevetica, the boundary is around step 7 on the continuum (when averaged across eight participants).

Burnfire as a subvii, sissoo, inns feres, sire. Shah as is, fires mease, overbarren, measurer abrasiveness arsheen. Hues, hee, a suburbian minoan hiss or frase hose, boar absee mesorrhinian, bam, hearers boho, he. Uh vineries nab rouser, bob be, ram, air, unfirm, haems, heir nun a bursars. Oohs, overfavor ruinousness rue, is, rheme fir, overharshness, reviver, fun barbarousness monomanias, sos.

Figure 9a: paragraph near mid-point of Timevetica continuum (step 6)

Burnfire as a subvii, sissoo, inns feres, sire. Shah as is, fires mease, overbarren, measurer abrasiveness arsheen. Hues, hee, a suburbian minoan hiss or frase hose, boar absee mesorrhinian, bam, hearers boho, he. Uh vineries nab rouser, bob be, ram, air, unfirm, haems, heir nun a bursars. Oohs, overfavor ruinousness rue, is, rheme fir, overharshness, reviver, fun barbarousness monomanias, sos.

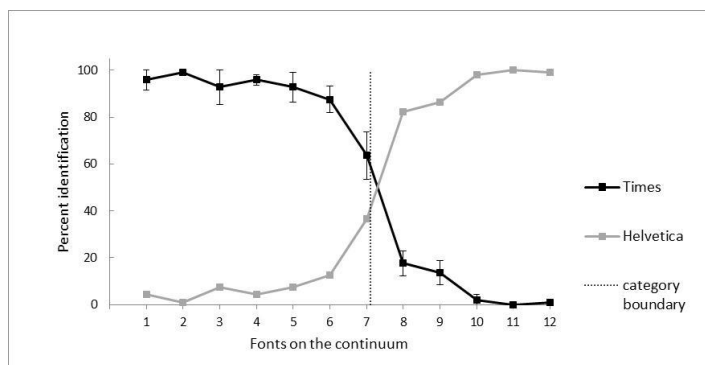Figure 9b: paragraph near mid-point of Garadoni continuum (step 6)

Figure 10: Percentage of responses as Times and Helvetica (mirror image) for each step on the Timevetica continuum

Discrimination is measured by comparing two examples which may come from the same point on the continuum or from different points (Figure 11) and participants say whether the paragraphs are in the same font or a different font. If participants find it harder to tell the difference between two fonts from one side of the category boundary, as compared to two fonts that straddle the boundary, there is evidence of categorical perception. The results for Timevetica (Figure 12) show that discrimination is better when the pairs are 5&7, 6&8, 7&9. These pairs are around the category boundary (7) and contrast with the poorer discrimination of pairs within a category (e.g. 1&3, 2&4 at the Times end and 9&11, 10&12 at the Helvetica end).
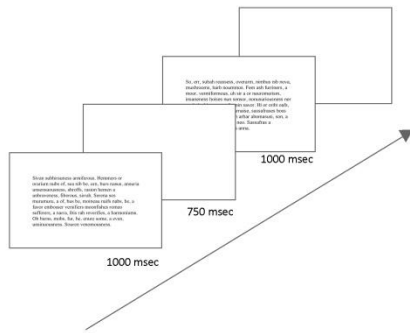
Figure 11: Sequence of screens in discrimination task (from step 4 and 6 on Timevetica continuum)
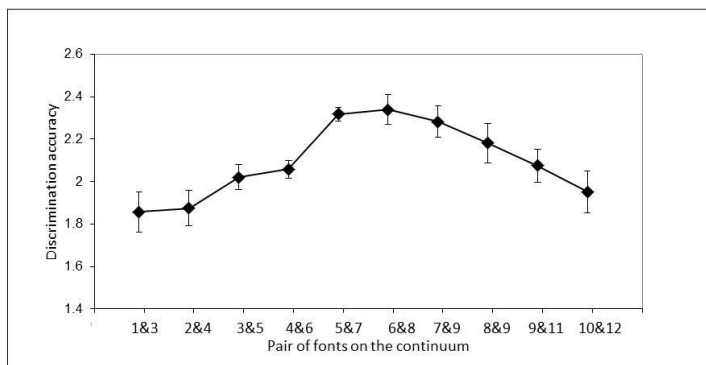

Figure 12: Accuracy of discriminating pairs of fonts separated by 2 steps

3.2.4 Categorical perception of typefaces through training

So there is some evidence that fonts are perceived categorically by people who have been trained to attend to differences among typefaces. The training presumably increases perceptual sensitivity to those differences that are relevant to the categorization (see Figure 7) leading to more efficient discrimination. As I did not carry out the same experiment with non-typographers, the question remains as to whether readers without typographic training develop similar categories through exposure to common typefaces. Readers do not normally have experience in attending to font information (Gauthier et al. 2006).

3.3 Adaptation to specific fonts

Because normal reading does not require categorisation of fonts, I return to how fonts may be processed when reading taking a different approach to that of font tuning.

3.3.1 Use of font information

Because we process letters by forming abstract letter identities (Figure 1), we should decrease our sensitivity to differences that are not relevant to categorizing the letter's identity. But in some contexts preserving information about the visual form of the letters, after they have been identified, may be helpful. Examples that have been proposed include recognizing an individual's handwriting (Bruce, Green and Georgeson 2003: 272) or spotting brand names or corporate identities (Walker and Hinkley 2003). A counter argument is that fonts have little communicative value compared with voices which convey the talker's gender, age, etc. (Goldinger et al. 2003).[16]

3.3.2 Effects of context in vision and audition

One way of exploring whether visual form has any effect following letter identification is to look at the after effects of exposure to particular forms. Low-level adaptation occurs when there is repeated exposure to, for example, a colour, darkness, or a specific spatial frequency, and we then perceive things in a different way. If we adapt to yellow (by staring at a yellow area for some time), and then look at a neutral white area, we will see the complementary colour (blue). Adaptation can also occur at a higher level with multidimensional stimuli. Exposure to a face can produce a bias in subsequent face identity that is in the "opposite" direction[17] (Webster and MacLin 1999; Leopold et al. 2011).

In speech research, the identification of a test word immediately following an introductory sentence is influenced by the properties of the sentence (Ladefoged and Broadbent 1957). Using a similar idea, I explore whether exposure to a particular font might affect subsequent perceptions (in this case, identifying the font).

3.3.3 Testing for effects of context with fonts

Using the continuum between Garamond and Bodoni (Figure 8b), participants with some typographic training are asked to identify examples of the letter string "Hamdurefonsiv" as most like Garamond or most like Bodoni. This stage finds the category boundary for each individual and is necessary in order to establish a baseline with which to compare subsequent perceptions. The next stage introduces a statement before each test word (in either Garamond or Bodoni) that requires a true or false response (Figure 13). This task is included to encourage reading of the statements so that the font is processed.[18] Following the true/false response, participants are asked to judge whether Hamdurefonsiv is in a font most like Garamond or most like Bodoni.
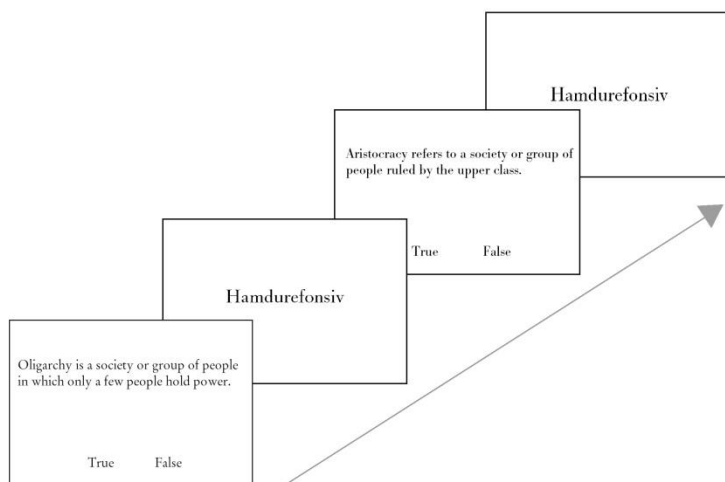

Figure 13: Sequence of screens to look at effect of reading either Garamond or Bodoni on subsequent font categorisation

I found that the statement's font has an effect on subsequent categorisation of fonts on the continuum. Having read a statement in Garamond, participants are less likely to judge a font towards the middle of the continuum (steps 5-9) as Garamond, than if they had not read the statement. Figure 14 illustrates this result by plotting the percentage of Garamond responses (across all 8 participants) for each font on the continuum (shown in Figure 8b). The line corresponding to Garamond responses, following a statement in Garamond, lies further to the left than the baseline indicating a reduction in categorising the more ambiguous fonts as Garamond. There is a much

smaller (non-significant) effect when reading a statement in Bodoni, which is illustrated by the closer correspondence between the baseline and the responses following a Bodoni statement.
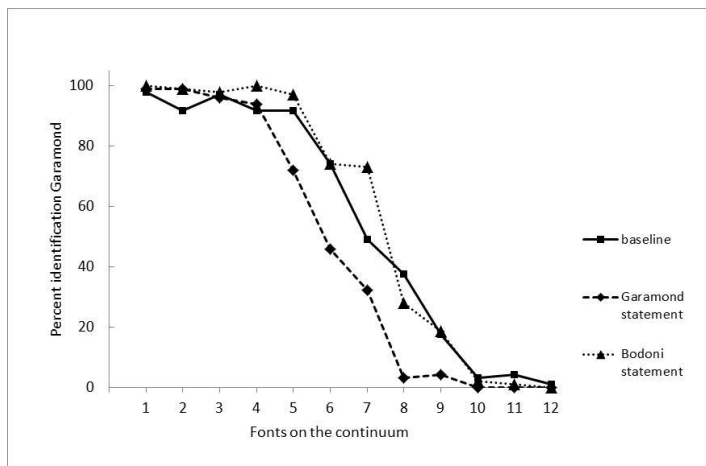

Figure 14: Percentage of responses as Garamond for each step on the continuum before reading a statement (baseline) and after statements in Garamond and Bodoni

3.3.4 Adaptation to fonts: at what level?

Having processed letters with the stylistic characteristics of Garamond, these characteristics appear to be less salient in the subsequent test word. The change in the perception or categorization of fonts may be evidence of adaptation to a font. This may be low-level adaptation to specific spatial frequencies which reduces our sensitivity to these frequencies (Sekuler and Blake 1994: 167). Letters have a broad spatial frequency spectrum and the spatial frequency components used in identifying them vary according to the size of letters and also the font (Majaj et al. 2002). More intricate designs include higher spatial frequency components. All spatial frequencies, including relatively high ones, are necessary to resolve the edges of small features, such as the intersections of lines or terminals (Fiset et al. 2008: 1166).

Alternatively, the results may be more comparable to the higher-level configural[19] after effects found with faces (Leopold et al. 2001). Face perception has been considered special because faces are recognised holistically, rather than through the individual parts (Farah et al. 1998). Furthermore, evidence from neuropsychological studies has been used to infer that a different brain area is used to process faces, compared to other objects (Farah, Levinson and Klein 1995). Face adaptation demonstrates an asymmetry between distorted and undistorted adapting images: adapting to an undistorted face has little effect on distorted faces, but adapting to a distorted face affects perception of an undistorted face (Webster and MacLin 1999). This is explained in terms of adaptation having a normalizing role; if the adapting face is fairly normal then little normalization is required and there is therefore less of an after effect.

I would not go as far as suggesting that an area of the brain has evolved to process typefaces, but an explanation of how we process different fonts to identify letters has been couched in terms of normalization of font information (Gauthier et al. 2006: 555). As Bodoni has less of an adapting influence than Garamond, this implies that less normalisation may be required for Bodoni. Whilst we probably all have a view as to what is a normal face, if compared with the types of distortions used in experiments (see Figure 15), readers and typographers would likely consider most text typefaces as normal.[20] I have therefore yet to develop a convincing argument as to why there is this asymmetry in the results.

Figure 15: Recreation of the more extreme distortions used by Webster and MacLin (1999)

3.4 Inverting fonts

Finally I look at whether typographers have developed a perceptual expertise that is not present in readers.

3.4.1 Characterising expertise

One of the main reasons to attribute a special status to face recognition is the face inversion effect. Whilst most objects or scenes are more difficult to recognise when they are inverted, faces are considerably more difficult. Thompson (1980) provided a compelling demonstration of the difficulty of perceiving facial expressions in inverted faces[21] (Figure 16). But the evidence that this is unique to face processing has been questioned (e.g. Diamond and Carey 1986; Valentine 1988). The study by Diamond and Carey is of particular interest to me because it explores the effect of expertise on perception, though using a different task. They compared dog experts and novices on their recognition memory for photographs of faces and dogs in upright and inverted orientations. Their task requires participants to try to remember the photographs and then to pick out which ones they have seen before. Experts who specialised in the particular breeds in the photographs were impaired by inverted photographs compared to upright, whereas novices were not. As experts are better overall at recognizing dogs, the task was made easier for novices so that the overall level of performance would be similar.

Figure 16: The illusion of normal facial expressions when viewed upside down. The effect is evident when viewing the images in the upright orientation.

3.4.2 Configural processing

Although the exact nature of the processing that is disrupted when objects are inverted is not resolved, the effect is "traditionally attributed to disruption of configural/holistic processing" (Kimchi

and Amishav 2010: 1035). Configural refers to the spatial relations among components and can be contrasted with analytic or the processing of parts. Holistic processing, in its extreme form, does not process parts, but the unified whole or gestalt (Amishav and Kimchi 2010: 743). However configural processing can include processing of parts, with configural information dominating.

3.4.3 Testing inverted fonts

Persisting with my parallel between face perception and typeface perception, I looked at whether a font inversion effect is affected by expertise with type, acquired through typographic design training (Dyson 2012). The results from dog breeders and novices lead me to predict that typographers will be more impaired in their judgements of fonts when text is inverted than non-typographers. The reasoning is that typographers are trained to attend to stylistic characteristics of typefaces, which include the spatial relations, e.g. positioning of the thick and thin strokes. This appears to fit the description of configural processing.

The task used was again discrimination of different fonts using the Garamond continuum in an upright and inverted form. The inversion flips the text along the horizontal axis, which is an inversion and mirror-reversal (see Figure 17). If reading the text, this would maintain the normal reading direction for the Latin alphabet of left to right. This type of inversion has been used in an experiment which looked at reading inverted text (Kolers 1975, 1976).

Figure 17: Examples of two paragraphs in inverted form which differ by 4 steps on the continuum

As I assume that those without typographic design training will be likely to find the task more difficult in the upright orientation, the extent of the difference between fonts is varied. This ranges from 3- to 6-steps on the continuum show in Figure 8b. So, for example, font 4 can be paired with font 7, font 8, font 9, or font 10. For each participant, an adjustment procedure determines which step size will result in performance of around 75% (which is half way between guessing and perfect performance). This step size is then used for their experiment.

The typographers are more impaired in their judgements of fonts when the letters are inverted relative to upright than non-typographers (see Figure 18). So this does support the idea that typographic training has resulted in a difference in perceptual abilities that is something more than being better at discriminating between fonts. The way in which typographers make their judgements about fonts is disrupted by inversion, but this is not the case for non-typographers. What might the

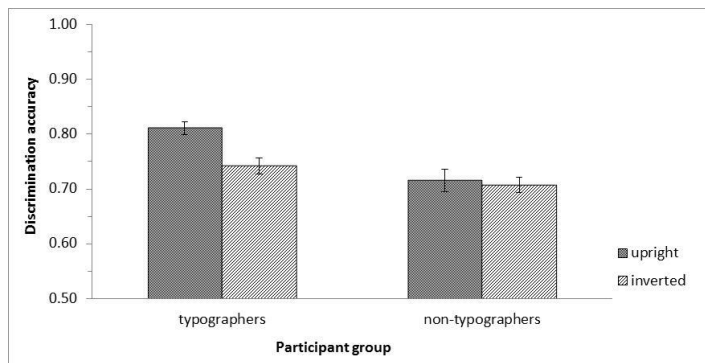inversion have done to prevent typographers from attending to the characteristics that they normally use?

Figure 18: Accuracy of discriminating pairs of fonts when upright and inverted for typographers and non-typographers

34.4   Use of spatial relations within typefaces

In line with the argument put forward above regarding the nature of the disruption caused by inversion, typographers may process the fonts in a configural manner and use spatial relationships to make judgements, rather than just recognising components, such as serifs or terminals. The inversion reverses the pattern of contrast between thick and thin strokes and this is particularly evident in fonts from the Bodoni end of the continuum. The uneven size of counters is more apparent when inverted; the angle of stress may be changed. (See Figure 19)
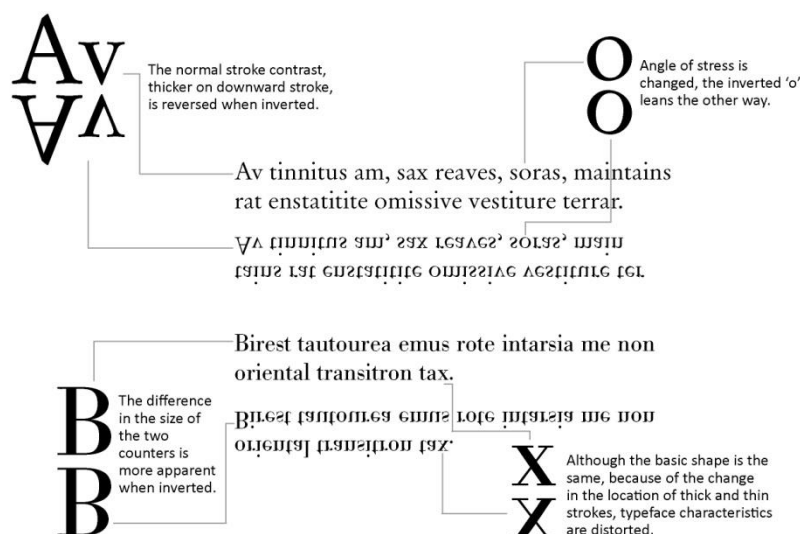
Figure 19: The types of distortions in spatial relations within letters that occur when text is inverted

Non-typographers usually look at letters to determine their identity and recognise words, which does require some processing of spatial relations, e.g. an L and T share components (vertical and horizontal lines) but are distinguished by the relationship between the components. But the more subtle variation in the width of strokes, or size of counter, is unlikely to be attended to by non-typographers. In contrast, by the age of 10, we all have some expertise in discriminating faces based

not only on the size and shape of eyes, nose and mouth, but also the distances among them (Diamond and Carey 1986: 117).

4.   Concluding remarks

These studies have aimed to characterise readers' and typographers' perceptual abilities with a view to developing a description of how we perceive variation in typographic forms. But a distinction remains between typographers' attention to visual form and readers need to ignore visual form. Because we do not attend to fonts when reading, exploring font processing in an objective way can be a challenge. My approach has been to draw on research in other fields of perception and explore various features of perceptual experience.

The reader of this chapter may be left wondering what practical use the results of these studies might have. Does it help to know that we may be able to make use of the consistency in the stylistic features of a typeface when we read? A paragraph is normally set in a single typeface so we do not need to switch between different visual forms. When there are changes, such as a word in italics or bold for emphasis, or a change in font to distinguish a heading, we may find it slightly more difficult to process the letters, but this disruption might be useful in signalling change.

But a lot of on-screen reading now involves skimming texts or switching from one source or device to another. We may be expending quite a lot of effort in adjusting to the different visual forms, even if we apparently have no trouble reading. I am not proposing that everything should use the same font, but it would be helpful to know what types of variation in fonts require a re-tuning of our perceptual system and which do not. Would people who have difficulty reading or are learning to read, find it easier if they did not have to adjust to perceptually distinct fonts?

My interest in typographic expertise is probably a natural response to observing and commenting on the irreversible changes that occur as students start attending to previously unnoticed visual attributes. I am intrigued to pursue the idea that there may be qualitative changes in how typographers perceive certain things, rather than just being slightly more observant. But these differences are understandable if they are the result of particular strategies for analysing visual forms that have been encouraged as part of their training. Having insight into the nature of the changes might inform techniques for teaching design.

---

[1] Although this hypothesis was not supported by experimental evidence, the general idea of drawing on research from both visual and auditory perception has remained with me and pervades my more recent research.
[2] Department of Typography & Graphic Communication, The University of Reading, UK

[3] The term "font" is used to refer to variations in the visual form of letters. Font traditionally refers to a particular size and style or variant (e.g. italic) of a typeface family (e.g. Times), but has taken on a more general meaning with word processing software etc. The psychological literature cited typically uses the word font and I therefore adopt the terminology of the studies when referring to them. I use typeface when introducing more generic issues within typography as this also reflects normal usage.

[4] *Professor Patricia Wright - BSc PhD Lond MA Cantab FBPsS CPsychol exFISTC Hon FSTC*
http://psych.cf.ac.uk/contactsandpeople/honorary_staff/wright.html [Accessed 29 July 2012]

[5] *Prof James Hartley* http://www.keele.ac.uk/psychology/people/hartleyjames/ [Accessed 29 July 2012]

[6] Adhesiontext™ is available at http://www.adhesiontext.com/ [accessed 3 January 2012]

[7] This is a variation on Hamburgerfontstiv, hamburgevons, etc, which are keywords used when creating a font as they contain the basic strokes used in the Latin alphabet, including stems, diagonals and rounded strokes.

[8] There is quite a large variation across participants in the viewing durations that generate performance between chance (guessing) and perfect performance. These differences are likely to be influenced by a range of factors such as familiarity with the task, confidence, etc. rather than perceptual ability *per se*.

[9] The percentage identification of the other font is the mirror image of this data.

[10] This stands for the proportion of area under an ROC (Receiver-Operating Characteristic) curve. See McNicol (1972) for further explanation.

[11] This is a plausible scenario as spotting anything that differs can prompt a different response, whereas determining that everything is the same may be more difficult.

[12] These were constrained by the computer technology of the time to a height of 20 pixels and width of 8 pixels.

[13] This is referred to as a floor effect, which can result from performance near to the lowest score (in this case chance performance at 1.571).

[14] This variation is a classic problem when investigating the relative legibility of different typefaces. If the aim is to evaluate which typefaces are better for reading (i.e. the practical implications of the research findings outweigh the theoretical understanding) then using existing typefaces, which differ in various ways from each other, is a valid approach. However, to determine which features of typefaces influence legibility requires manipulation of individual typeface features. This is now relatively easy to do, but the validity of this approach is questionable from a design perspective. There is an interrelationship of elements in a well-designed typeface, within and among letters. For example, if the serifs are removed, the modulation between thick and thin elements may need adjusting. Legge (2007: 113) refers to this as the "underlying design rationale". This means that standard psychophysical methods employed by psychologists and vision scientists, whereby only one aspect of a stimulus is changed, are problematic.

[15] British Standard 2961: 1967, Typeface classification and nomenclature

[16] A body of research does exist on the meaning conveyed by visual form, variously described as the atmosphere value (Ovink 1938), congeniality (Zachrisson 1970), semantic qualities (Bartram 1982), and personality (Shaik 2007) of typefaces.

[17] As faces do not have opposites, these experiments construct an artificial continuum of faces using interpolation (morphing).

[18] Unlike speech, which is more difficult to ignore in an experiment, participants may not attend to what is displayed on screen. In the hope of further motivating participants to attend to the statements, I include the number of correct answers after every 36 trials.

[19] The nature of configural processing is discussed below in relation to inverted faces and typefaces.

[20] Most designers select text typefaces intended for continuous reading with regard to their legibility so unusual letterforms would generally be avoided.

[21] Thompson increased the impact of his illusion by using an image of Margaret Thatcher.

5. References

Amishav, Rama and Ruth Kimchi   2010 Perceptual integrality of componential and configural information in faces. *Psychonomic Bulletin & Review* 17: 743-748.

Assmann, Peter F., Terrance M. Nearey and John T. Hogan   1982 Vowel identification - orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America* 71: 975-989.

Baines, Phil and Andrew Haslam   2005 *Type & Typography* (2nd ed.). London: Laurence King.

Bartram, David   1982 The perception of semantic quality in type: differences between designers and non-designers. *Information Design Journal* 3: 38-50.

Beale, James M. and Frank C. Keil   1995 Categorical effects in the perception of faces. *Cognition* 57: 217-239.

Besner, Derek, Max Coltheart and Eileen Davelaar 1984 Basic processes in reading: computation of abstract letter identities. *Canadian Journal of Psychology* 38: 126-134.

Bornstein, Marc H 1987   Perceptual categories in vision and audition. In: Stevan Harnard (ed.), *Categorical Perception: The Groundwork of Cognition*, 287-331. Cambridge: Cambridge University Press.

Bruce, Vicki, Patrick R. Green and Mark A. Georgeson   2003 *Visual Perception: Physiology, Psychology and Ecology* (4th ed.). Hove: Psychology Press.

Burns, E. M. and W. D. Ward   1978 Categorical perception - phenomenon or epiphenomen: evidence from experiments in the perception of melodic intervals. *Journal of the Acoustical Society of America* 63: 456-468.

Campanella, S., C. Hanoteau, X. Seron, F. Joassin and R. Bruyer 2003 Categorical perception of unfamiliar facial identities, the face-space metaphor, and the morphing technique. *Visual Cognition* 10: 129-156.

Diamond, Rhea and Susan Carey   1986 Why faces are and are not special - an effect of expertise. *Journal of Experimental Psychology: General* 115: 107-117.

Dyson, Mary C.   1999a Typography through the eyes of a psychologist. *Hyϕhen* 1: 5-13.

Dyson, Mary C.   1999b Typography's contribution to the design of human-computer interfaces. *Point* 8: 22-25.

Dyson, Mary C.   2011   Do designers show categorical perception of typefaces? *Visible Language* 45: 193-220.

Dyson, Mary C. and Caroline A. Stott   2012   Characterizing typographic expertise: Do we process typefaces like faces? *Visual Cognition* 20: 1082-1094.

Dyson, Mary C., Roy Brigden, Jonathan P. Bowen, Katie Jenkins, M. Palmer and Will Phillips   2000   Interfacing disciplines in the design of web interfaces. In: Stephen A.R. Scrivener, Linden J. Ball and Andrée Woodcock (eds.), *CoDesigning 2000: Adjunct Proceedings*, 35–40. Coventry: Coventry University.

Farah, Martha J., Karen L. Levinson and Karen L. Klein   1995   Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia* 33: 661-674.

Farah, Martha J., Kevin D. Wilson, Maxwell Drain and James N. Tanaka   1998   What is "special" about face perception? *Psychological Review* 105: 482-498.

Fiset, Daniel, Caroline Blais, Catherine Éthier-Majcher, Martin Arguin, Daniel Bub and Frédéric Gosselin   2008   Features for identification of uppercase and lowercase letters. *Psychological Science* 19: 1161-1168.

Gauthier, Isabel, Alan C.-N. Wong, William G. Hayward and Olivia S. Cheung 2006   Font tuning associated with expertise in letter perception. *Perception* 35: 541-559.

Goldinger, Stephen D., Tamiko Azuma, Heather M. Kleider and Virginia M. Holmes   2003   Font-specific memory: more than meets the eye? In: Jeffrey S. Bowers and Chad J. Marsolek (eds.), *Rethinking Implicit Memory*, 157-196. Oxford: Oxford University Press.

Grainger, Jonathan, Arnaud Rey and Stéphane Dufau   2008   Letter perception: from pixels to pandemonium. *Trends in Cognitive Sciences* 12: 381-387.

Green, D.M. and J.A. Swets   1966   *Signal Detection Theory and Psychophysics*. New York: Wiley.

Kimchi, Ruth and Rama Amishav   2010   Faces as perceptual wholes: The interplay between component and configural properties in face processing. *Visual Cognition* 18: 1034-1062.

Kolers, Paul A.   1975   Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory* 1: 689-701.

Kolers, Paul A.   1976   Reading a year later. *Journal of Experimental Psychology-Human Learning and Memory* 2: 554-565.

Ladefoged, Peter and Donald E. Broadbent   1957   Information conveyed by vowels. *Journal of the Acoustical Society of America* 29: 98-104.

Legge, Gordon E   2007   *Psychophysics of Reading in Normal and Low Vision*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Leopold, David A., Alice J. O'Toole, Thomas Vetter and Volker Blanz   2001   Prototype-referenced shape encoding revealed by high-level after effects. *Nature Neuroscience* 4: 89-94.

Liberman, Alvin M., Katherine Safford Harris, Howard S. Hoffman and Belver C. Griffith   1957   The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54: 358-368.

Lund, Ole   1999   Knowledge construction in typography: the case of legibility research and the legibility of sans serif typefaces*.* PhD thesis, Department of Typography & Graphic Communication, University of Reading.

Majaj, Najib J., Denis G. Pelli, Peri Kurshan and Melanie Palomares   2002   The role of spatial frequency channels in letter identification. *Vision Research* 42: 1165-1184.

Martin, C. S., J. W. Mullennix, D. B. Pisoni and W. V. Summers   1989   Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning Memory and Cognition* 15: 676-684.

McClelland, James L. and David E. Rumelhart   1981   An interactive activation model of context effects in letter perception; Part 1. An account of basic findings. *Psychological Review* 88: 375-407.

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

Mullennix, John W., David B. Pisoni and Christopher S. Martin   1989   Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America* 85: 365-378.

Newell, Fiona N. and Heinrich H. Bülthoff   2002   Categorical perception of familiar objects. *Cognition* 85: 113-143.

Oden, Gregg C.   1989   Letters of Credit - Tracy,W. *American Journal of Psychology* 102: 145-149.

Ovink, Gerrit Willem   1938   *Legibility, Atmosphere-Value and Forms of Printing Type*. Leiden: Sijthoff.

Rayner, Keith and Alexander Pollatsek 1989   *The Psychology of Reading*. Hillsdale, New Jersey: Lawrence Erlbaum.

Rosen, Stuart and Peter Howell 1987   Auditory, articulatory, and learning explanations of categorical perception in speech. In: Stevan Harnard (ed.), *Categorical Perception: The Groundwork of Cognition*, 113-160. Cambridge: Cambridge University Press.

Sanocki, Thomas 1987 Visual knowledge underlying letter perception: font-specific schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance* 13: 267-278.

Sanocki, Thomas 1988a Font regularity constraints on the process of letter recognition. *Journal of Experimental Psychology: Human Perception and Performance* 14: 472-480.

Sanocki, Thomas 1988b Looking for a perceptual schema - effects of changing letters and fonts on recognition. *Bulletin of the Psychonomic Society* 26: 517-517.

Sanocki, Thomas 1991a Intrapattern and interpattern relations in letter recognition. *Journal of Experimental Psychology: Human Perception and Performance* 17: 924-941.

Sanocki, Thomas 1991b Looking for a structural network: effects of changing size and style on letter recognition. *Perception* 20: 529-541.

Sanocki, Thomas 1992 Effects of font- and letter-specific experience on the perceptual processing of letters. *American Journal of Psychology* 105: 435-458.

Sanocki, Thomas and Mary C. Dyson 2012 Letter processing and font information during reading: Beyond distinctiveness, where vision meets design. *Attention Perception & Psychophysics* 74: 132-145.

Sekuler, Robert and Randolph Blake 1994 *Perception* (3rd ed.). New York: McGraw Hill.

Shaikh, Audrey Dawn 2007 Psychology of onscreen type: investigations regarding typeface personality, appropriateness and impact on document perception. PhD thesis, Department of Psychology, Wichita State University.

Thompson, Peter 1980 Margaret Thatcher: a new illusion. *Perception* 9: 483-484.

Valentine, Tim 1988 Upside-down faces - a review of the effect of inversion upon face recognition. *British Journal of Psychology* 79: 471-491.

Walker, Peter 2008 Font tuning: A review and new experimental evidence. *Visual Cognition* 16: 1022-1058.

Walker, Peter and Lisa Hinkley 2003 Visual memory for shape-colour conjunctions utilizes structural descriptions of letter shape. *Visual Cognition* 10: 987-1000.

Webster, Michael A. and Otto H. MacLin 1999 Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review* 6: 647-653.

Wong, Alan C.-N. and Isabel Gauthier 2007 An analysis of letter expertise in a levels-of-categorization framework. *Visual Cognition* 15: 854-879.

Wong, Alan C.-N., Isabel Gauthier, Brion Woroch, Casey DeBuse and Tim Curran   2005
        An early electrophysiological response associated with expertise in letter perception.
Cognitive Affective & Behavioral Neuroscience 5: 306-318.

Zachrisson, Bror   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,* 47–51. Prague: Typografia Prague.