Information Design Journal 22(1), **3–18**

DOI: 10.175/idj.22.1.02dys

Mary Dyson & Sofie Beier

Investigating typographic differentiation

Italics are more subtle than bold for emphasis

Keywords: typeface, typography, legibility, salience

Text designers are likely to benefit from guidance on how to use typographic differentiation for emphasis. Three experiments use purposely-designed fonts to explore the size and nature of differences in the stylistic characteristics of fonts (weight, width, contrast, italic) which affect letter identification. Results indicate that words set in bold and expanded fonts, when alternated with words set in a Neutral test font, may impair performance, whereas changing to italic does not. Possible explanations are explored through measuring the physical and perceptual similarities of the test fonts.

1. Introduction

In his influential book *The Elements of Typographic Style* Robert Bringhurst recommends that bold fonts can be used "to flag items in a list, to set titles and subheads u&lc in small sizes, to mark the opening of the text on a complex page, or to thicken the texture of line that will be printed in pale ink…" (Bringhurst 1992: 56). He goes on to recommend that if sparingly used, bold fonts can effectively emphasise keywords and numbers in a dictionary. In addition to the bold weights, the italic style is often used for emphasis, mainly for words or sentences within longer paragraphs of running text set in roman. The present study looks into how these, and other typographical features, when combined with a regular font, influence our ability to rapidly identify words.

How is knowledge of reading processes relevant to design practice? The way in which the psychologist Legge (2007: 107–8) defines legibility provides the link. Although the visual properties of text (e.g., font, size) influence legibility, the outcome is primarily determined by characteristics of visual processing, the reader's perceptual representation of the text. There is therefore an argument for providing designers with a reasonable understanding of the mechanisms of reading so they may anticipate how readers will process their work. However, the research questions addressed by reading researchers may not be of direct relevance to the practice of designers. The mechanisms underlying skilled reading are researched by psycholinguists (Rastle 2007). However, their interest tends to start with abstract letter identities (Figure 1), which refer to the identification of 'a' or 'b', regardless of what the letter looks like (i.e., font, case, size and colour). Those who design fonts or select fonts (e.g., typographers; graphic, book, information, instructional, and sign designers) have a particular interest in the precise visual form.

A less mainstream approach to letter perception, which incorporates a design perspective, has looked at

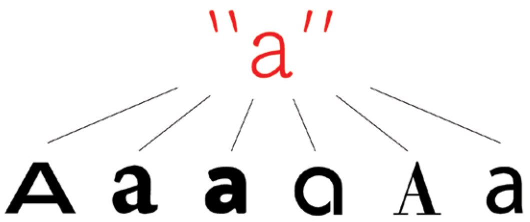


Figure 1. Different visual forms of letter 'a' which all map onto a single abstract letter identity.

commonalities among letters within a font which can support letter identification. Letters in the same font share stylistic characteristics and this uniformity of design has long been recognised in type design practice. The original research was pioneered by a psychologist, Sanocki (1987, 1988, 1991a, 1991b) who explained how the particular characteristics of a font can contribute to more efficient letter identification. Sanocki (1987) proposed a model whereby the perceptual system becomes tuned to a particular font over time ('font-tuning') and within this model a set of font parameters describe the visual aspects of fonts. These parameters are used to transform font-specific input (precise visual form) into font-invariant form (abstract letter identity). Sanocki demonstrated that letters presented in a single font were easier to identify than letters in mixed fonts, described as a 'regularity effect', i.e., advantage for single fonts.

Determining what type of variation and how much variation is tolerated before letter identification becomes more difficult has practical relevance in text design, where italic or bold distinguishes a word from the surrounding text. In such contexts, letter identification may be marginally slowed down but this may not be deleterious. Words in bold may also be perceptually more salient, described by Tinker (1965) as having greater visibility. The degree of overlap between letters in different fonts (a measure of physical differences) has been investigated but no clear correlation found between amount of overlap and accuracy of letter identification (Sanocki 1991b).

A relatively small number of studies have demonstrated the regularity effect, usually based on faster reaction times (RT) to a single font, compared with mixed fonts (Sanocki 1987, 1988; Gauthier, Wong, Hayward, & Cheung 2006; Walker 2008). These studies used diverse experimental conditions, with material ranging from dot matrix letters, consistent with the computer technology of the time (Sanocki 1987, 1988) to current but very dissimilar typefaces, Cooper Black and Palatino Italic (Walker 2008).

The above studies all investigated the regularity effect in terms of letter identification, except for Walker (2008) who used a lexical decision task. This involves determining whether a string of letters is a word or a non-word (described as a pseudoword when one or two letters are changed). This task therefore involves word recognition, whereas other studies (Sanocki 1987; Dyson 2014) used unrelated letter strings (i.e., not forming a word or pseudoword) or single letters (Gauthier et al. 2006). The current experiments explore whether regularity effects can be found in the context of words. This closer approximation to normal reading, though still not meaningful sentences, is a move towards the demands of designers, and some psychologists, for ecological validity.

In our studies we use Rapid Serial Visual Presentation (RSVP) to display words on screen, where single words are presented sequentially in the same position. This technique allows us to adjust the rate of presentation. Typically reading is faster with RSVP (Legge 2007), which may induce errors in word recognition. Employing a lexical decision task with RSVP is intended to increase the sensitivity of the experimental task and the likelihood of detecting differences in accuracy or response times due to visual forms. Typographic variation tends to produce very small differences in performance and the skills of fluent readers can obscure effects.

The research has two aims: to further explore reading mechanisms through scientifically rigorous experimentation and to inform typographers and type designers who are concerned with creating texts that are easy to read. Three experiments investigate the presence of a regularity effect in words using a set of fonts that vary systematically from a Neutral test font. All fonts were designed for the experiments. The objective is to determine which stylistic features might produce a regularity effect and to further explore whether the incidence and extent of the effect is dependent on the size or nature of the visual and perceptual differences between fonts.

While the previous experiments into font tuning have compared fonts of different typefaces (Sanocki 1988; Walker 2008), or have compared fonts that are highly different from a normal reading situation (Gauthier et al. 2006), the present investigation applies test fonts that can be described as a typeface family, as they are related through having a common appearance. When spatial relationships are adjusted (e.g., letter width in relation to x-height) normal relationships between parts of letters are still maintained. This enables us to isolate stylistic features, and test fonts that might be combined in normal reading material.

2. General method

2.1 Participants

Participants in the experiments were volunteer university students who were reimbursed for their time. Twelve people participated in each of the experiments; no one participated in more than one experiment. Design students were not included to ensure that the results reflect the processes normally employed when reading, rather than sensitivity to typeface characteristics, which might result from training in observing the details of typefaces. The research project was reviewed according to the procedures specified by the University Research Ethics Committee and was given a favourable ethical opinion for conduct.

2.2 Apparatus and test material

The experiments were run on a Dell Latitude D820 laptop with a TFT-LCD of 8.7″ by 13.56″ and a diagonal of 15.4″, set to a resolution of 1280 × 800 (96dpi). E-Prime software controlled the timing and presentation of material, recorded responses, and provided feedback on the accuracy of responses.

The aim in creating the test material was to use fonts of a professional look equal to what a reader will meet in his or her everyday life. A master font named Neutral Regular was designed, inspired by an idea put forward by the type designer Frutiger (1998). Frutiger suggests that all characters are modelled around a basic letter shape, also called 'letter skeleton'. This skeleton shows itself as the dark area that emerges when the most popular typefaces are superimposed. Neutral Regular is based on the superimposed fonts: Garamond Regular (Monotype), Monotype Baskerville Regular (Monotype), Bodoni Book (Adobe), Excelsior Roman (Adobe), Times Regular (Linotype), Palatino Roman (Adobe), Optima Roman (Linotype), and Univers 55 Roman (Linotype). With the letter strokes placed at the darkest areas of the superimposed typefaces, the aim is for Neutral Regular to embody the notion of a generic letter (Figure 2).

We created 6 fonts based on the master font, Neutral Regular, described in Beier (2013). Three of these fonts have one stylistic feature added, and three have two stylistic features added. The stylistic features are weight, width, contrast, and italic (Figure 3). All fonts have the same x-height, as different x-heights are known to affect the perceived size of a font (Beier 2012).

Figure 2. Top left: Eight superimposed fonts; Bottom left: Neutral Regular—based on the darkest area of the superimposed fonts.

1	Neutral Regular (the master font)
	the quick brown fox jumps over a lazy dog
2	Neutral Italic (one extra stylistic feature: italic)
	the quick brown fox jumps over a lazy dog
3	Neutral Bold (one extra stylistic feature: weight)
	the quick brown fox jumps over a lazy dog
4	Neutral Expanded (one extra stylistic feature: width)
	the quick brown fox jumps over a lazy dog
5	Neutral BoldItalic (two extra stylistic features: weight + italic)
	the quick brown fox jumps over a lazy dog
6	Neutral Contrast (two extra stylistic features: weight + contrast)
	the quick brown fox jumps over a lazy dog
7	Neutral BoldExpanded (two extra stylistic features: weight + width)
	the quick brown fox jumps over a lazy dog

Figure 3. Master font and 6 variations differing by one or two stylistic features.

The stylistic feature named 'italic' differs from the other stylistic features by being based on a number of different elements. All the letters of the italic stylistic feature are slanted 15° and condensed by 90%. As the italic letter style originates in the cursive writing hand, the letter skeleton of italic letters such as 'a' and 'g' differ substantially from the same letters in the master font (Figure 4). The combined influence of these elements causes all the italic letters to have a structure that is different from the letters of the master font.

Figure 4. Left: Neutral Italic superimposed on Neutral Regular demonstrating the different letter skeleton. Right: Neutral Bold and Neutral Regular superimposed show identical letter skeleton.

2.3 Procedure

All experiments used a within-subject design, so that each participant experienced all font conditions for that experiment. The font conditions were number of fonts in a trial (single font versus two fonts), and stylistic features (font pairs) within the experiment. In all cases, one of the two fonts in a font pair was the Neutral Regular font (Figure 3, number 1) and stylistic features were examined by pairing this with each of the six variations (Figure 3, numbers 2–7). The choice of font pairs for each experiment was determined to some degree by the results of the previous experiments, as well as the need to include all six variations.

Each trial within an experiment consisted of the sequential presentation of 15 six- or seven-letter words in 24 point, with an x-height of 4mm, in the centre of the screen. The trial began with a fixation cross in the centre of the screen displayed for 300 ms, followed by the words left aligned with the central fixation point, providing a consistent horizontal location for the first letter (Legge 2007), and a 700 ms interval between trials (Figure 5).

Participants viewed the screen from around 55 cm. The x-height was approximately 0.42 degrees. The length of words varied from about 17 mm (e.g., the word 'little') to 32 mm ('common') in Neutral Regular, corresponding to 25 mm and 54 mm in the Neutral Bold Expanded font.

In half the trials, the eleventh or the twelfth item was a non-word (as in Figure 5: 'tandin') and all other items were words; the other half of trials contained all words. The target non-word/word was placed in these positions to enable sufficient words to have been read prior to the target. Walker (2008) suggests that there may be unconscious strategic control of font information, and incentives may be required to remain tuned to a font. Repeated use of the same font may provide the incentive.

Having been presented with 15 items, participants were required to indicate whether they detected a non-word or all words in a two-alternative forced-choice task. The scale of responses ranged from 'sure non-word' to 'non-word' to 'words' to 'sure words' corresponding to 4 keys on the keyboard. This scale allowed the participant to adopt different criteria reflecting confidence in their decision. They were given feedback on their response with either a green tick or a red cross. The words were randomly selected from a set of 300 high frequency words (between 80 and 2000 per million), obtained from the MCW Orthographic Word Form database (Medler & Binder 2005). The database also generated 300 non-words of the same length with unconstrained trigram-based strings to make them less word-like.

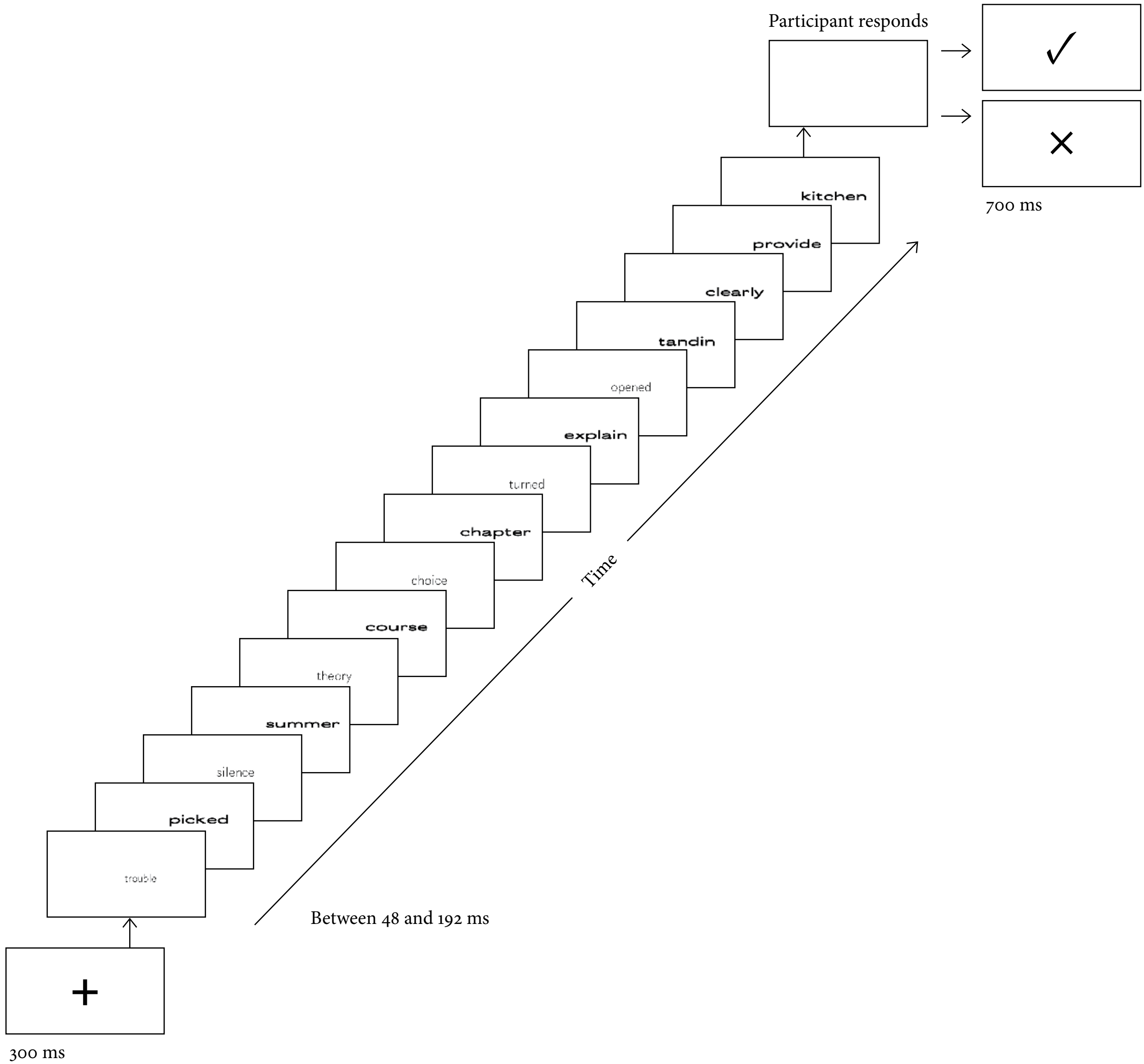


Figure 5. Sequence of screens in one trial with two fonts (Regular and Bold Expanded) and including a non-word.

Fonts were mixed by presenting alternate words in the two fonts; the word following the target item and the remaining three or four words of the trial were presented in the same font as the target, i.e., alternation stopped from position 12 or 13 to the end (Figure 5). Making the final part of the trial equivalent for single and two-font trials ensured that any effects of number of fonts was due to the first 11 or 12 words.

To equate the level of difficulty across participants as far as possible, a staircase adjustment procedure was carried out prior to each of the experiments. This aimed to identify the word duration for each participant that resulted in a performance level of around 75%, i.e., midway between chance (guessing) and perfect performance. This adjustment block of 80 trials used the same procedure as the main experiment and started with a duration of 208 ms for each word, staying the same, or adjusting up or down in multiples of 16 ms according to the number of correct responses in the preceding 4 trials. At the end of this block, the duration was set and remained constant through the main experiment blocks. Across the three experiments, the individual participant durations of words ranged from 48 ms to 192 ms; the slower reading rates were therefore around 5 words per second.

2.4 Data analysis

Two measures were used in analysing the data: reaction times and accuracy of responses. Reaction time was measured from the onset of the blank screen following the last word (Figure 5: Participant responds) to the key press. A discrimination index p(A) was calculated to determine accuracy, rather than simply counting the number of correct responses. This index incorporates the confidence level (e.g., sure non-word versus non-word).

Differences between single and two fonts, and among font pairs, were analysed for each measure using Analysis of Variance (ANOVA). These differences were explored in two ways: (i) bundling together the two single font conditions within a font pair; (ii) separating out each of the single fonts. The second approach was used as words presented in each font within a pair may differ in their efficiency of recognition (accuracy or RT) suggesting different levels of legibility. Although measuring legibility is not a primary objective of the studies, such results can inform design practice. The data from targets presented in a particular font was compared in the context of the single font and when alternating with a second font (two font condition). This analysis may also reveal differences in perceptual salience. For example, in font pairs where the two font condition alternates words presented in a heavier font with words presented in a lighter font, the target may be detected more easily when it is set in the heavier font through being more perceptually salient.

Where appropriate, reported probability values used the Huynh-Feldt correction following tests for sphericity. All graphs show mean results across 12 participants and include within-subject standard error bars, calculated using Cousineau-Morey corrections (O'Brien & Cousineau 2014). These reflect the difference between conditions and control for the variance between participants.

3. Experiment 1

The first experiment used two font pairs: Neutral Regular with Neutral Expanded, and Neutral Regular with Neutral Bold Expanded, comparing one additional stylistic feature (width) with two (weight and width).

3.1 Method

The two font pairs were presented in separate parts and each part was preceded by an adjustment block for that pairing. Having established duration, each part consisted of 4 blocks of 48 trials with breaks in between. Within each block, trials were equally divided between single font and two fonts, and two font trials included equal numbers starting with each of the two fonts. The order of trials was randomised within each block for each participant. Half the participants received Neutral Regular with Neutral Expanded in the first part of the experiment and then moved on to Neutral Regular and Neutral Bold Expanded. The other half followed the reverse order.

3.2 Results

A within-subject ANOVA found differences in RT for the number of fonts ($F(1,11) = 13.57$, $p = 0.004$) with slower responses after words were presented in two fonts than in a single font. Words in the two font pairs had similar RTs and although there appeared to be a larger difference between single and two fonts with two additional stylistic features (weight and width) than width alone (Figure 6) this result was not statistically significant.

When the two single fonts were combined in the analysis there were no differences in discrimination (accuracy) scores between single fonts and two fonts and no difference between the two font pairs. However, separating the data according to the fonts of targets found a different pattern of results in discrimination scores for Neutral Regular compared with Neutral Expanded (Figure 7), a significant interaction between font (of target) and number of fonts ($F(1,11) = 4.98$, $p = 0.047$). When the target word was presented in Neutral Regular, accuracy was impaired if preceded by a word presented in Neutral Expanded (i.e., two fonts). This was not the case when the target was set in the Neutral Expanded font where accuracy was similar in the context of single or two alternating fonts. This analysis revealed no statistically significant results for the other font pair: Neutral Regular and Neutral Bold Expanded.

The difference in discrimination of the two single fonts, seen in Figure 7, indicates that words in Neutral Expanded were recognised less well than Neutral Regular words, suggesting that Neutral Expanded is less legible.

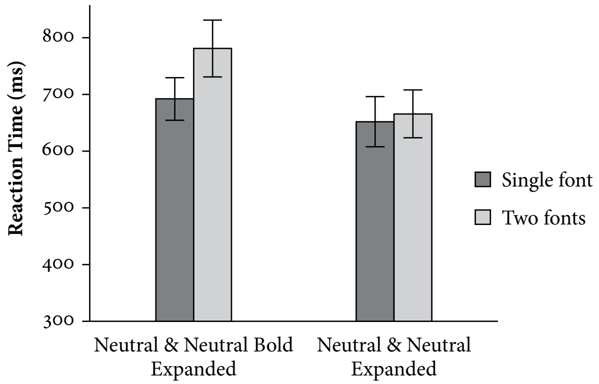


Figure 6. Experiment 1: Mean reaction times by font pair (Neutral Regular with Neutral Bold Expanded and Neutral Expanded) comparing number of fonts.

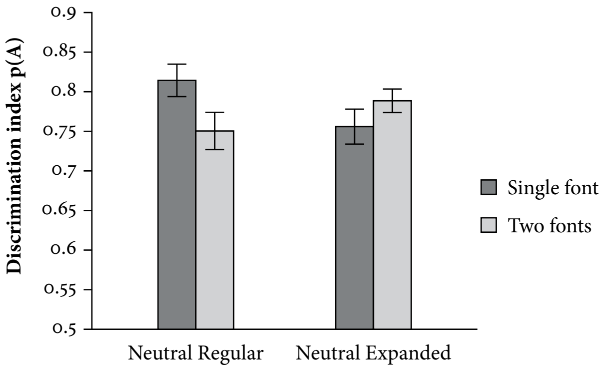


Figure 7. Experiment 1: Discrimination accuracy of Neutral Regular and Neutral Expanded targets in the context of single fonts or two alternating fonts.

4. Experiment 2

This experiment explored whether weight affects performance in combination with other stylistic features. Three font pairs were compared: Neutral Regular with Neutral Bold Expanded, Neutral Bold Italic, and Neutral Contrast. All pairs vary in two stylistic features, one of which is weight. Including Neutral Bold Expanded provided some continuity with the previous experiment and enabled comparison with the other features; Neutral Bold Italic added a change to the skeleton; Neutral Contrast added stroke contrast.

4.1 Method

To accommodate three pairings, this experiment randomised font pairs within each block, rather than separating into parts. A single adjustment block was followed by 10 blocks, each containing 40 trials, with 16 single font and 24 two font trials. When broken down into font pairs, this ratio maintained equal numbers of single and two font trials, of which half contained non-words.

4.2 Results

Figure 8 shows that RTs varied according to font pair ($F(2,22) = 5.69$, $p = 0.022$) with words in Neutral Regular and Neutral Bold Expanded responded to slower than the other two pairs. The effect of number of fonts was not statistically significant ($F(1,11) = 3.72$, $p = 0.08$); responses to trials with words set in two fonts were not reliably slower than those set in a single font.

These results show a much weaker effect of number of fonts than Experiment 1 (smaller differences in RT). Response times to words in the font pair Neutral Regular with Neutral Bold Expanded were slower than other font pairs across single and two fonts. The single font data (left bar in each pair in Figure 8), although combining two single font RTs does suggest that words in Neutral Bold Expanded may be less legible than the other two fonts (Neutral Contrast and Neutral Bold Italic), as Neutral Regular RT is a constant across font pairs. There were no significant differences in discrimination scores.

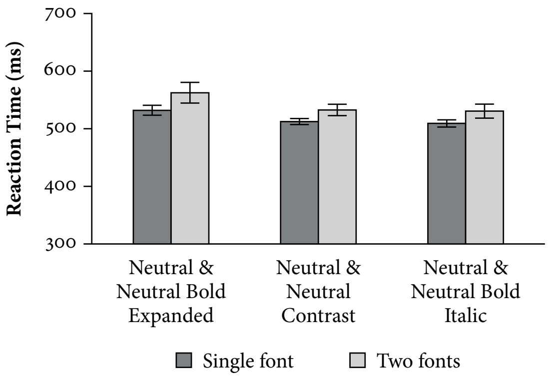


Figure 8. Experiment 2: Mean reaction times of three font pairs (Neutral Regular with Neutral Bold Expanded, Neutral Contrast and Neutral Bold Italic) comparing number of fonts.

5. Experiment 3

The remaining two font pairings (Neutral Regular with Neutral Italic and Neutral Regular with Neutral Bold) were tested in the third experiment, with the addition of Neutral Bold Italic. Pairing with Neutral Bold provided a means of establishing whether weight on its own is sufficient to produce a regularity effect, by comparison with Neutral Bold Italic; pairing with Neutral Italic further explored whether weight is necessary, i.e., does a change that does not involve weight but instead a change to the skeleton, result in a regularity effect.

5.1 Method

The method was the same as Experiment 2, i.e., randomizations of all variables within blocks.

5.2 Results

There was no significant difference in RT across font pairs but words set in two fonts were generally responded to slower than a single font ($F(1,11) = 5.98$, $p = 0.033$). This was not entirely consistent across font pairs (Figure 9). The significant interaction between font pair and number of fonts, ($F(2,22) = 4.47$, $p = 0.027$) indicated that in two font pairs (Neutral Regular and Neutral Bold; Neutral Regular and Neutral Bold Italic) switching between fonts slowed responses; weight alone was sufficient to produce a regularity effect. However, switching between words set in regular and italic did not slow down responses.

Analysing the accuracy of words set in each font within a font pair, Neutral and Neutral Italic revealed no further differences. However, there was a statistically significant interaction between target font and number of fonts ($F(1,11) = 6.15$, $p = 0.03$) in discrimination of words presented in Neutral Regular and Neutral Bold (Figure 10). In this pairing, the accuracy of discriminating words presented in each font was similar when presented in single font trials, suggesting similar levels of legibility. However, the insertion of words set in Neutral Bold tended to decrease accuracy of targets set in Neutral Regular, whereas words set in Neutral Regular did not decrease accuracy of detecting targets set in Neutral Bold.

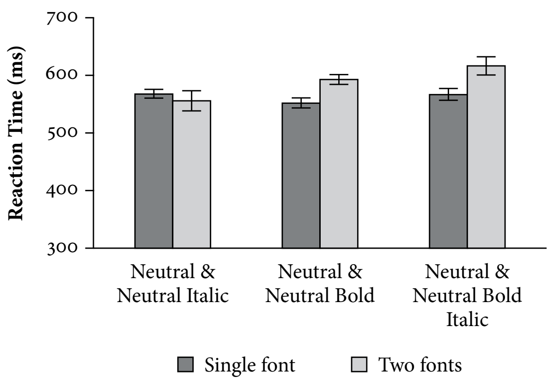


Figure 9. Experiment 3: Mean reaction times of three font pairs (Neutral Regular with Neutral Italic, Neutral Bold and Neutral Bold Italic) comparing number of fonts.

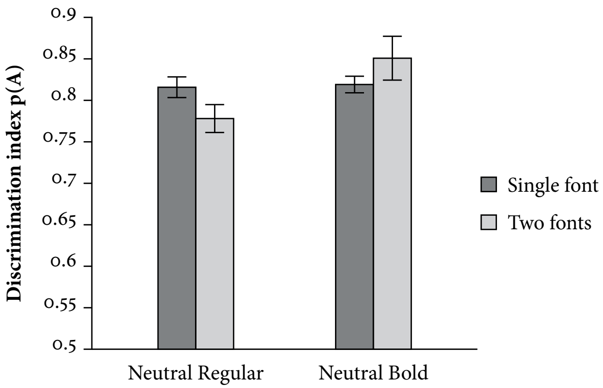


Figure 10. Experiment 3: Discrimination accuracy of Neutral Regular and Neutral Bold targets in the context of single fonts or two alternating fonts.

This asymmetry was not found with Neutral Regular and Neutral Bold Italic. However, targets set in Neutral Bold Italic (in single and two font trials) were better discriminated than targets set in Neutral Regular ($F(1,11) = 6.85$, $p = 0.024$). This suggests that Neutral Bold Italic is a more legible font than Neutral Regular (Figure 11).

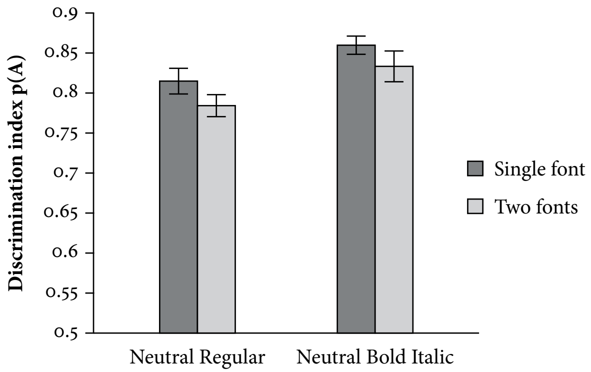


Figure 11. Experiment 3: Discrimination accuracy of Neutral Regular and Neutral Bold Italic targets in the context of single fonts or two alternating fonts.

5.3 Discussion

The results of these three experiments suggest:

- the addition of weight may induce regularity effects, evident from words presented in single fonts being processed more quickly than words presented in two fonts
- Neutral Expanded is less legible than Neutral Regular, and this also seems to be true of Neutral Bold Expanded; Neutral Bold Italic is more legible than Neutral Regular
- mixing Neutral Regular with either Neutral Expanded or Neutral Bold decreases the accuracy of recognizing targets set in Neutral Regular but facilitates to a small extent, or does not impair, recognition of targets set in the other font

Although these results indicate some differences consistent with regularity effects which may be influenced by the degree of overlap between fonts, they also suggest effects due to some stylistic features being more salient than others. To explore these explanations, two methods were used to gauge the similarities between Neutral Regular and the other fonts, independent of the RSVP data. Physical similarities were measured by computing overlap; perceptual similarities were determined by a ranking procedure. Quantifying perceived similarities between Neutral Regular and the other fonts may confirm which fonts are relatively more salient.

6. Physical similarities: Overlap computation

6.1 Method

This method was intended to gauge the degree of difference between fonts independent of the nature of the difference, to examine the results in relation to the overlap hypothesis (Sanocki 1991b). To quantify physical overlap, each font was compared with Neutral Regular, letter-by-letter, shifting horizontally to the position maximizing overlap. Pixel correlations were computed for each letter, weighted according to the letter frequency in the sample of words and non-words used in the RSVP experiments, to produce an average for the six fonts.

6.2 Results

The average weighted correlations of 26 letters between Neutral Regular and all other fonts are illustrated in Figure 12. The letters of the two italic fonts were rotated, so that the italics and regular had similar angles on the stems. The values in the left column indicate that Neutral Contrast is most similar to Neutral Regular, while Neutral Bold Expanded is least similar. The two expanded fonts have far less overlap, due to the width of letters; by this measure they may be considered the most physically dissimilar. Words set in Neutral Expanded also reduced the accuracy of discriminating a target set in Neutral Regular. However the reduction in accuracy from Neutral Bold does not appear to be due to the degree of overlap, as this font has far more overlap than other fonts which do not have the same effect on discrimination.

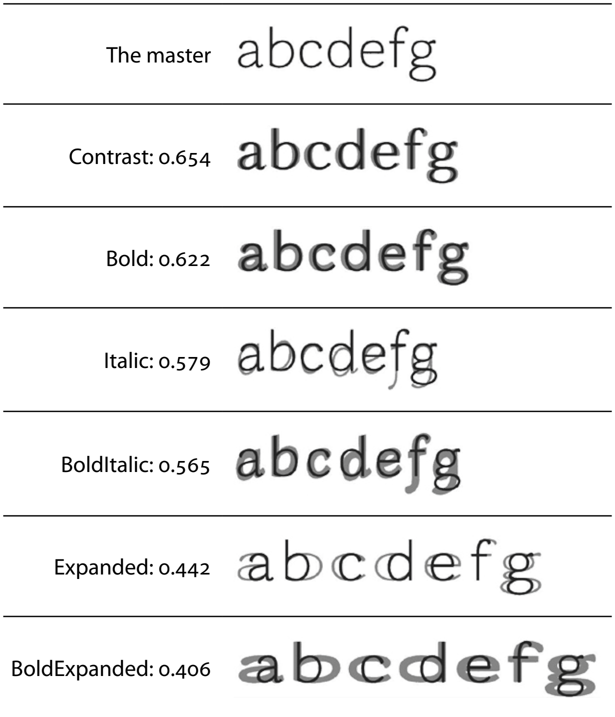


Figure 12. Correlations between the master font Neutral Regular and the 6 other fonts. The letters of the two italic fonts were rotated, so that the italics and regular have similar angles on the stems. The values in the left column indicate that Neutral Contrast is most similar to Neutral Regular, while Neutral Bold Expanded is least similar.

7. Perceived similarities: Ranking procedure

7.1 Method

Two groups of design students were recruited: one comprised seventeen final year undergraduate students studying Graphic Communication within the Department of Typography & Graphic Communication at the University of Reading, UK; the other comprised seventeen undergraduate and graduate students from The Royal Danish Academy of Fine Arts. These results were compared with rankings from the non-designer participants who took part in Experiments 1–3, who completed this procedure after the RSVP task. Thirty five results were obtained as one participant said they were unable to make the judgements.

Participants were shown example sentences (the quick brown fox jumps over a lazy dog) set in Neutral Regular and the six other fonts on separate cards and asked to rank the fonts in their similarity to Neutral, where 1 = most similar and 6 = least similar.

7.2 Results

Rankings were averaged within the group of designers and non-designers and are illustrated in Figure 13. Kendall's Coefficient of Concordance indicated that there was agreement within groups though less agreement among designers ($W = 0.276$, $\chi^2 = 46.866$, $p < 0.005$) than non-designers ($W = 0.389$, $\chi^2 = 68.126$, $p < 0.005$). Informal comments from designers suggested that they had difficulty in deciding whether bold, italic or expanded was the most significant feature. Non-designers may also have faced this dilemma, but not voiced their concerns. Although the order varied slightly between the two groups, there was general agreement ($r = 0.93$, $p = 0.008$). Neutral Bold Expanded, Neutral

Bold Italic, and Neutral Contrast were perceived as less similar to Neutral Regular, all of which have two stylistic features added to the Neutral Regular font.

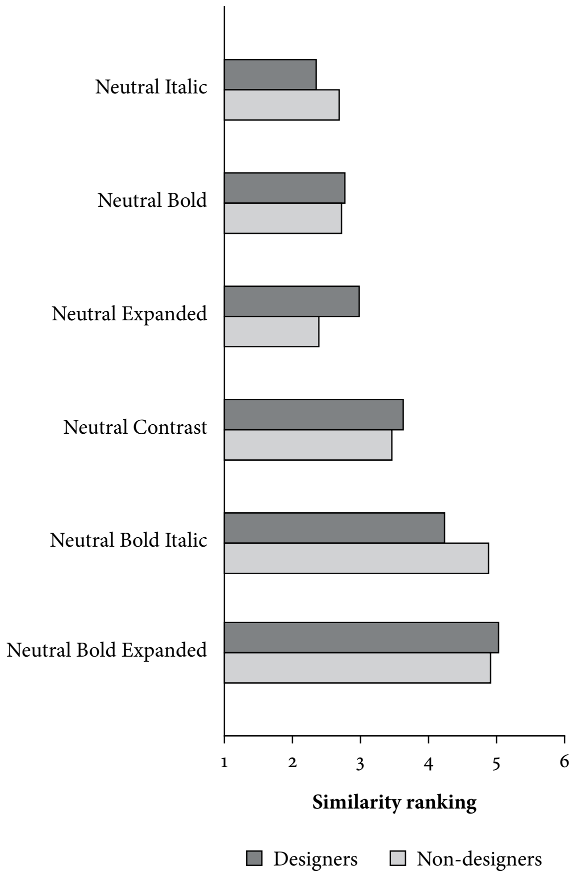


Figure 13. Average ranks for designers and non-designers of similarity to Neutral Regular (1 = most similar; 6 = least similar).

8. Comparing physical and perceptual similarities

Participants' rankings of similarities are very weakly based on the overlap between letters, as there is a non-significant correlation between the two measures ($r = 0.358$, $p = 0.48$). Two fonts which differed in physical and perceived similarity are Neutral Contrast and Neutral Expanded. Participants judged Neutral Contrast (varying in weight and stroke contrast) as fairly different from Neutral Regular, whereas this font has the most pixel overlap. A possible reason for this disjunction may the distribution of non-overlapping pixels, i.e., where they are located in relation to the skeleton (Neutral Regular). Varying stroke width in Neutral Contrast adjusts the rhythm of the strokes. This may have a more substantial effect in perceptual terms than more non-overlapping pixels where these are fairly evenly distributed around the skeleton (i.e., Neutral Bold). Also even though Neutral Contrast has additional weight, it may be perceived as lighter than Neutral Bold, i.e. more similar to Neutral Regular, because it has light hairline strokes (Beier 2013). Conversely, widening letters without additional weight (Neutral Expanded) was not perceived as quite as dissimilar as the lack of overlap suggests. In making similarity judgements non-designers in particular would appear to be adjusting the horizontal scale of the letters, to a certain extent, demonstrating a form of shape and size constancy (Haber & Hershenson 1980).

The convergence on Neutral Bold Expanded as the most dissimilar to Neutral Regular provides some validation of their contrast, suggesting words set in Neutral Bold Expanded may be perceptually salient in the contexts of alternating with Neutral Regular. However, Neutral Bold has considerable pixel overlap and is perceived as quite similar to Neutral Regular, which questions whether perceptual salience provides a sufficient account of the RSVP results.

9. General discussion

The results indicate that we are slower to respond to words which are presented in two fonts which vary in either one or two stylistic features. An exception is the mix of Neutral Regular and Neutral Italic where RT is unaffected. As Neutral Italic is judged as most similar to Neutral, this may account for the efficiency of switching between these two fonts, and would suggest that words set in Neutral Italic do not stand out very much when mixed with Neutral Regular. Italic is also a font that is conventionally used within texts to emphasise single words, and readers may therefore be more familiar with this mixture of fonts than other combinations. Across experiments, there is a suggestion that two stylistic features produce larger changes in RT, and the similarity data confirms that these fonts are less similar to Neutral Regular.

Analysis of discrimination accuracy suggests that the stylistic features affecting performance are width or weight. As these fonts vary in only one feature, the nature of the change is important and not just the number of changes. Gauthier et al. (2006) also found only one of their font manipulations produced a regularity effect, where the x-height of letters was abnormally large compared to fairly small. There is only limited support for overlap as an explanation. Although the Neutral Expanded font has a small physical overlap with Neutral Regular, the much larger degree of overlap between Neutral Regular and Neutral Bold argues against the extent of physical differences as the sole determinant of disruptions to accuracy of discrimination.

The results highlight a distinction between the legibility of a single font and the perceptual saliency acquired by the font when compared with Neutral Regular. Tinker (1965: 136) argues for a similar distinction in relation to bold, proposing that it "should not be employed for printing entire pages or books" but can be used for contrast and emphasis. This difference may help to explain the asymmetry in regularity effects.

The expanded fonts are not particularly legible as single fonts. If legibility can be expressed in terms of the ease with which we can tune to a font, Neutral Regular would be a candidate for ease of tuning. It embodies the basic letter shape which should require little transformation to an abstract (font-invariant) representation. Neutral Bold Expanded and Neutral Expanded fonts, by changing the spatial parameters and being physically distinct, may challenge the font tuning mechanism, resulting in reduced legibility.

Neutral Bold, on the other hand, may be tuned to more easily, as spatial parameters require no adjusting and bold is commonly mixed with regular fonts, e.g., for headings. To explain the regularity effect when targets set in Neutral Regular are mixed with Neutral Bold, we assume that Neutral Bold is also perceptually salient. Informal observation of Figure 3 would suggest that this is indeed the case, even though similarity measures do not support this. Neutral Bold Italic and Neutral Contrast do not show a similar effect in discrimination (Experiments 2 & 3); italic and contrast may interfere with weight to reduce the salience, rather than adding distinctiveness.

The findings are of interest to information designers by providing evidence of the perceptual status of ways to highlight text and indications of the relative legibility of words presented using RSVP. For centuries, typesetters and designers have employed italic and bold fonts for emphasis. The present investigation confirms the efficacy of this mix using current digital fonts, but demonstrates a distinction between the two. Summarising these results in the form of guidelines for the use of stylistic features:

– words set in a neutral/regular style are a good basis for efficient letter recognition, whereas expanded fonts hinder legibility

– setting words in italic can distinguish text elements without significantly disrupting reading and is therefore suitable for inclusion within continuous text
– bold is more appropriate than italic for setting headings or other access devices through making words stand out
– a more subtle distinction may be achieved through adding to a bold style to create a Bold Italic or (Bold) Contrast

As these results come from using an RSVP method, a variant of which is incorporated into some current reading technologies, there is also guidance as to how changes in typography might affect reading from small screen devices.

Submission date: 6 Nov, 2015
Accepted date: 29 Feb, 2016

Acknowledgements

We thank Kevin Larson for generating letter correlations; Clare Leake, Francesca Romano, Stéphane Passerat, and Matthew Standage for research assistance.

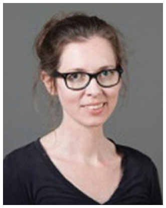
References

Beier, S. (2012). *Reading letters: designing for legibility.* Amsterdam: BIS Publishers.
Beier, S. (2013). Legibility investigations: controlling typeface variables. *Praxis and Poetics: Research through Design, Conference Proceedings* (pp. 92–95). Newcastle upon Tyne and Gateshead.
Bringhurst, R. (1992). *The elements of typographic style*. Point Roberts, WA: Hartley & Marks Publishers.
Dyson, M.C. (2014). Applying psychological theory to typography: is how we perceive letterforms special? In D. Machin (Ed.), *Visual Communication* (pp. 215–242). Berlin: De Gruyter Mouton.
Frutiger, A. (1998). *Signs and symbols: their design and meaning.* London: Ebury Press.
Gauthier, I., Wong, A.C.-N., Hayward, W.G., & Cheung, O.S. (2006). Font tuning associated with expertise in letter perception. *Perception, 35*, 541–559. DOI: 10.1068/p5313
Haber, R.N., & Hershenson, M. (1980). *The psychology of visual perception*. New York, London: Holt, Rinehart and Winston.
Legge, G.E. (2007). *Psychophysics of reading in normal and low vision.* Mahwah, New Jersey: Lawrence Erlbaum Associates.
Medler, D.A., & Binder, J.R. (2005). MCWord: An On-Line Orthographic Database of the English Language. http://www.neuro.mcw.edu/mcword/
O'Brien, F., & Cousineau, D. (2014). Representing error bars in within-subject designs in typical software packages. *The Quantitative Methods for Psychology, 10*(1), 56–67.
Rastle, K. (2007). Visual word recognition. In M.G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 71–88). Oxford: Oxford University Press.
Sanocki, T. (1987). Visual knowledge underlying letter perception: font-specific schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance, 13*(2), 267–278. 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. DOI: 10.1037/0096-1523.14.3.472
Sanocki, T. (1991a). Intrapattern and interpattern relations in letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 924–941. DOI: 10.1037/0096-1523.17.4.924
Sanocki, T. (1991b). Looking for a structural network: effects of changing size and style on letter recognition. *Perception, 20*(4), 529–541. DOI: 10.1068/p200529
Tinker, M.A. (1965). *Bases for effective reading.* Minneapolis: Lund Press.
Walker, P. (2008). Font tuning: A review and new experimental evidence. *Visual Cognition, 16*(8), 1022–1058. DOI: 10.1080/13506280701535924

About the authors

Mary Dyson is associate professor in the Department of Typography & Graphic Communication at the University of Reading, UK. She studied experimental psychology leading to a PhD in perception, and teaches theoretical and empirical approaches to typography and graphic communication. Her most recent research interests are driven by a desire to bridge the gap between scientists and designers and find commonalities. Her experiments have looked at issues such as how designers perceive typefaces compared to non-designers. This paper describes the outcomes of collaborative research with Sofie Beier.

Email: m.c.dyson@reading.ac.uk



Sofie Beier is a type designer and associate professor employed at the School of Design under The Royal Danish Academy of Fine Arts, where she is the head of the MA programme in Type & Wayfinding. She holds a PhD from the Royal College of Art in London and is the author of the book "Reading Letters: designing for legibility". Her current research is focused on improving the reading experience by achieving a better understanding of how different typefaces and letter shapes can influence the way we read.

Email: sbe@kadk.dk