Psicothema

Psicothema 2013, Vol. 25, No. 1, 13-17
doi: 10.7334/psicothema2012.141

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Why does the APA recommend the use of serif fonts?

Manuel Perea
Universitat de València

Abstract

Background: The publication norms of the American Psychological Association recommend the use of a serif font in the manuscripts (Times New Roman). However, there seems to be no well-substantiated reason why serif fonts would produce any advantage during letter/word processing. **Method:** This study presents an experiment in which sentences were presented either with a serif or sans serif font from the same family while participants' eye movements were monitored. **Results:** Results did not reveal any differences of type of font in eye movement measures –except for a minimal effect in the number of progressive saccades. **Conclusions:** There is no reason why the APA publication norms recommend the use of serif fonts other than uniformity in the elaboration/presentation of the manuscripts.

Keywords: reading, visual-word recognition, eye movements.

Resumen

¿Por qué la APA recomienda el uso de fuentes con serif? **Antecedentes:** las normas de publicación de la American Psychological Association recomiendan el uso de un tipo de letra con serif en los manuscritos (Times New Roman). Sin embargo, no parece haber ninguna razón bien sustentada por la que las fuentes con serif produzcan ventaja alguna durante la lectura de palabras. **Método:** para examinar el papel de los serifs se realizó un experimento en el que se presentaron frases bien con una fuente serif o sans serif de la misma familia, mientras se registraban los movimientos oculares de los participantes. **Resultados:** los resultados no revelaron diferencias debidas al tipo de letra en los movimientos oculares, a excepción de un efecto mínimo en el número de movimientos sacádicos progresivos. **Conclusiones:** no hay ninguna razón por la cual las normas de publicación de la APA recomienden el uso de otras fuentes con serif más allá de la uniformidad en la elaboración y presentación de los manuscritos.

Palabras clave: lectura, reconocimiento palabras, movimientos oculares.

Serifs are the small ornaments at the end of strokes which occur in many fonts (e.g., compare the X of a serif font with the X of a sans serif font). The publication norms of the American Psychological Association [APA] (2010) specify that manuscripts should use a serif font like Times New Roman (see also the Merriam-Webster's Manual for Writers and Editors, 2003)—and this includes the journal *Psicothema*. Indeed, the majority of books and e-books applications employ serif fonts. The present paper examines whether the use of a serif font provides an advantage over the use of a sans serif font during normal reading while the participants' eye movements are monitored—in terms of objective measures such as reading time, fixation duration, and number of saccades (see Tinker, 1963; Morrison & Inhoff, 1981, for a review of the early experiments of Tinker and others on typography and reading).

The choice between serif vs. sans serif fonts has generated lively debates among typographers in journals (e.g., Reynolds, 1979; Russell-Minda et al., 2007; Tinker & Paterson, 1932) and even movies (see Hustwit, 2007). In a key paper, Arditi and Cho (2005) examined two arguments which had been proposed in favor of serif fonts: i) serifs might increase letter discriminability by making the letter forms more complex, and ii) serifs might provide supplementary cues to the location of stroke ends. However, as Arditi and Cho argued, serifs are not an essential part of the letter, and one may consider these small ornaments as noise rather than signal. After all, why would the serifs in the letter M help the identification of its abstract letter representation in a computational/neural model of visual-word recognition (e.g., see Davis, 2011, for a computational model of visual-word recognition; see Dehaene, Cohen, Sigman, & Vinckier, 2008, for a hierarchical [neural-based] model of visual-word recognition)? Thus, there is no theoretically well-grounded model which favors that serifs facilitate (at some point) the process of letter/word recognition. Indeed, guide signs in roads, train/subway stations, museums, shopping malls, etc., across the world ordinarily employ sans serif fonts (e.g., Helvetica in the Chicago and New York subways). An excellent example is the Clearview font which was specifically designed to increase legibility in the distance—this is a *sans serif* font which is currently in use for guide signs in the US roads.

To examine the role of serifs during normal reading, the fonts under scrutiny should be essentially the same except for the use of serifs—that is, comparing Times New Roman (a serif font) vs. Helvetica (a sans serif font) would not be pertinent because these fonts differ in other potentially relevant parameters (see Sanocki & Dyson, 2012, for a recent review of typography and visual-word recognition). The published experimental studies comparing serif

Received: May 15, 2012 • Accepted: June 18, 2012
Corresponding author: Manuel Perea
Facultad de Psicología
Universitat de València
46010 Valencia (Spain)
e-mail: mperea@uv.es

vs. sans serif fonts are very scarce (Arditi & Cho, 2005; Moret-Tatay & Perea, 2011) and none of them employed a normal reading situation. Arditi and Cho (2005) created three typefaces that differed in the size of the serif (sans serif, as in b, intermediate, as in b, and large, as in b). In their experiments, participants had to read a rapid serial visual presentation of words or read a scrambled reading passage. They failed to obtain a significant effect of serif for the three participants in their study—one normally sighted and two participants with low-vision. Arditi and Cho concluded that the effects of serif are "unlikely to be of significance at typical print sizes viewed under normal conditions" (p. 2932). In addition, Moret-Tatay and Perea (2011) conducted a standard word identification experiment (lexical decision task: "is the stimulus a word?") comparing serif vs. sans serif fonts. In order to compare two fonts that essentially only differed in the presence (absence of serifs Moret-Tatay and Perea employed two fonts of the same family created by the same designer (Lucida vs. Lucida Sans; e.g., experiment vs. experiment). It is important to note here that both fonts are part of the same "typeface" (i.e., the overall design: the Lucida family; the term font a specific instantiation of a typeface [e.g., Lucida Sans 14pt, bold]). Moret-Tatay and Perea failed to find an advantage of the serif font—indeed, word-identification times were if anything faster with the sans serif font than with the serif font (the median size of the effect was only 6 ms). They concluded that "serifs do not seem to play a beneficial role in visual-word recognition—beyond being a decorative burden" (p. 623). It is important to note here that the use of the serif and sans serif version of fonts of the same family (Lucida) had already been employed in the unpublished rapid serial visual presentation experiment of Morris, Aquilante, Bigelow, and Yager (2002). Morris et al. failed to find an effect of serif in their experiment—except for a small deleterious effect of serifs when the font size was very small (a 4-pt size). One limitation of the Arditi and Cho (2005), the Moret-Tatay and Perea (2011) and the Morris et al. (2002) experiments is that these experiments did not employ a normal reading scenario and, thereby, it is unclear whether serifs provide an advantage in a normal reading setting. Furthermore, it is important to note that eye movement experiments can detect quite subtle effects during sentence reading. For instance, Slattery and Rayner (2010) found a small but significant advantage of using a sharper focus of the letters (via ClearType) relative to the default settings.

In sum, the goal of the present experiment was to examine whether or not the use of a serif font produces an advantage over a sans serif font in an ecological setting: normal reading. To that end, the participants' eye movements were monitored while reading one-line sentences (see Rayner, 2009; Rayner, Pollatsek, Ashby, & Clifton, 2012, for a detailed overview of the advantages of eye movement research). The sentences were written either with a serif font (Lucida) or with a sans serif font of the same family (Lucida Sans)—similarly to the word-recognition experiment of Moret-Tatay and Perea (2011) and the rapid serial visual presentation experiment of Morris et al., (2002). To examine the effect of serifs during normal reading, global measures were analyzed for each sentence (total time, number of saccades, or average duration fixation). In addition, local measures on a target word embedded in each sentences were also analyzed. More specifically, either a high- or a low-frequency target word was embedded in each sentence (see Slattery & Rayner, 2010, for a similar procedure). This way, it was possible to examine not only the local measures for a given word (e.g., first-fixation duration, gaze durations, and total time) but also to examine the potential interaction of the effect of serif with a relevant lexical factor such as word-frequency.

Method

Participants

Twenty-four undergraduate students from the University of Valencia took part in the experiment in exchange of a small monetary compensation (3€). All of them were native speakers of Spanish and had normal or corrected-to-normal vision. All the participants were naïve as to the purpose of the experiment. The study was conducted in accordance with the Declaration of Helsinki of 1975 (as revised in Tokyo in 2004).

Instruments

An Eyelink II eyetracker (SR Research Ltd, Canada) was employed to monitor the participants' eye movements. This is a video-based eye tracking device with a camera that samples pupil location at a rate of 500 Hz. Vision was binocular, but only the movements of the right eye were registered (see also Slattery & Rayner, 2010). The average gaze position error of the eyetracker is less than 0.5°. Participants were seated 60 cm from the computer screen. A head-tracking camera was used to compensate potential head motion.

A total of 120 experimental sentences were employed—these were the same as those used for Perea and Acha (2009). Each sentence included either a low-frequency word (60 words; mean frequency: 4.5 per million, number of letters: 7.3; Davis & Perea, 2005) or a high-frequency word (60 words; mean frequency: 87.3 per million; number of letters: 7.3). For the embedded target words, there were two sentence frames, as in "El niño tiró el cucurucho/ordenador al suelo delante de sus padres" and "La niña pidió un ordenador en su fiesta de cumpleaños" (cucurucho is a low-frequency word [cone, in English] and ordenador is a high-frequency word [computer, in English]). For each sentence frame, two lists of stimuli were created so that all participants read the 120 sentences with frequency and type of font being counterbalanced in a Latin Square design (i.e., 30 sans serif sentences with a low-frequency target word, 30 sans serif sentences with a high-frequency word, 30 serif sentences with a low-frequency target word, and 30 serif sentences with a high-frequency target word). The target words had a low predictability when embedded in the sentences (via a cloze task) and the sentences were simple to understand (see Perea & Acha, 2009, for further details on the materials).

Procedure

The experiment was carried out individually in a silent, dimly lit room. Participants were seated facing a computer screen and were told that in each trial, a sentence would appear on the computer screen. Participants were instructed to read the sentences for comprehension. Each trial began with a black square on the left side of the monitor. Once the participant looked at the square, the sentence appeared on a single line of text—the location of the square corresponded to the initial letter of the sentence. Participants were asked to press a key on a button game pad once they had finished reading the sentence. Before starting the experiment, the eyetracker was calibrated and the participant was asked to follow several points on the computer screen. The sentences were presented either in 14-pt Lucida or in 14-pt Lucida Sans. The 120 experimental sentences were preceded by 8 practice sentences to familiarize participants with the calibration procedure. Before starting each trial, calibration was checked—the eyetracker was recalibrated when necessary. The experimental sentences were presented in a different random order for each participant. Participants were asked yes/no comprehension questions after 20% of the sentences—they answered these questions with a high level of precision (accuracy= .96).

Data analyses

Both global measures and local measures were analyzed. The global measures were: i) total reading time (in ms), ii) average fixation duration following both progressive and regressive saccades (in ms), and iii) the number of progressive and regressive saccades. The only fixed factor for the global measures was type of font (serif, sans serif). Local measures on the target word were also examined—the fixed factors were type of font (serif, sans serif) and word-frequency (low, high): first fixation duration (i.e., the duration of the first fixation on the target word), gaze duration (i.e., the sum of the durations of all fixations on the target word before leaving it), and total time (i.e., the sum of the durations of all fixation durations on the target word—including both progressive and regressive saccades).

Results

The raw eye-tracking data were processed using EyeDoctor software from the UMass Community (http://www.psych.umass.edu/eyelab/software/). Fixations shorter than 80 ms that were within one letter of the following/previous fixation were merged into that fixation. In addition, to avoid the influence of extreme data, individual fixations shorter than 80 ms or longer than 800 ms were excluded (less than 3% of trials overall; see Slattery & Rayner, 2010, for a similar procedure). Eye fixation measures were then analyzed using the lme4 (i.e., linear mixed-effects) package in R. For each dependent variable, a succession of models of diminishing complexity of random effects structure was created (see Baayen, Davidson, & Bates, 2008). The fixed effects in the models were "font" (serif, sans serif) in the global analyses, and "font" (serif, sans serif) and "word-frequency" (low, high) in the local analyses. In all cases, the optimal model was the one that kept subjects and items as random effects. It is important to mention here that the "classical" by-participants ($F1$) and by-items ($F2$) analyses yielded exactly the same results as those reported here. The global and local eye fixation data—as well as the corresponding statistical tests— are shown in Tables 1 and 2, respectively.

Global analyses

The averages per type of font for all the dependent variables as well as the corresponding t-tests are displayed in Table 1. The results are clear: there are no signs of an effect of font, with the

Table 1
Global measures for each of the conditions: Total sentence reading time (in ms), progressive/regressive fixation duration (in ms), and number of progressive/regressive saccades

	Total reading time	Mean fixation duration		Number of saccades	
		Progressive	Regressive	Progressive	Regressive
Type of font					
Serif	2370	225	216	10.74	2.54
Sans Serif	2392 t= -1.001 SE= 21.93 p= .32	224 t= -0.88 SE= 1.22 p= .44	216 t= 0.08 SE= 3.06 p= .92	10.58 t= -2.26 SE= 0.06 p= .024 *	2.50 t= -0.45 SE= 0.06 p= .62

Table 2
Local measures (and the corresponding statistical analyses) for the different experimental conditions in the experiment: First fixation duration (in ms), gaze duration (in ms), and total time (in ms)

	First fixation duration		Gaze duration		Total time	
	Word-Frequency		Word-Frequency		Word-Frequency	
	Low	High	Low	High	Low	High
Type of font						
Serif	223	217	318	280	383	341
Sans serif	223	217	316	276	385	337
	Font: t= -0.01, p= .99 Frequency: t= 1.41, p= .15 Font × Frequency: t= 0.12, p= .91		Font: t= 0.35, p= .73 Frequency: t= 3.48, p= .0002* Font × Frequency: t= -0.12, p= .90		Font: t= 0.30, p= .77 Frequency: t= 2.84, p= .002 * Font × Frequency: t= -0.37, p= .70	

only exception of the number of progressive saccades, in which the sentences presented with a serif font produced (slightly) more fixations than the sentences presented in a sans serif font (10.74 vs. 10.58, respectively). However, this results must be observed with caution because—leaving aside the small magnitude of the effect— when the analysis included the overall number of saccades in the sentence (i.e., the sum of both progressive and regressive saccades), the effect of font was not significant, t= -1.53, p>.13.

Local analyses

As can be seen in Table 2, the local analyses on the target word did not show any trends of an effect of font—or an interaction with word-frequency. As usual, the statistical analyses revealed an effect of word-frequency on the target words—in particular for gaze durations and total time (39 and 45 ms, respectively; see Perea & Acha, 2009, for a similar pattern).

Discussion

Are serifs employed in books, manuscripts, etc., because of historical or aesthetical reasons rather than for scientifically-based evidence? As indicated in the Introduction, most publication style manuals specify that manuscripts, theses, etc. should be written using a serif font like Times New Roman. However, there are no well-founded theoretical reasons to use of a serif font over a sans serif font—beyond subjective preferences. Indeed, current models of visual-word recognition do not assign any role to the serifs in the process of letter/word recognition (e.g., spatial coding model, Davis, 2011). Furthermore, previous experiments failed to find a facilitative role of serifs in scrambled passage reading, rapid serial visual presentation, and word-identification tasks (see Arditi & Cho, 2005; Moret-Tatay & Perea, 2011; Morris et al., 2002). The data of the present eye movement experiment are clear and reveal that the presence/absence of serifs does not affect the process of normal reading—beyond a marginal effect in the number of progressive saccades. At the same time, the experiment obtained the usual pattern of word-frequency effects during the processing of the target words (i.e., the lack of an effect of serif was not due to lack of power or lack of sensitivity). There is a caveat, though. In the present experiment, all sentences were one-line sentences. One could argue that in order to entirely rule out the role of serifs during normal silent reading, it would be important to include conditions with multiple lines of text (i.e., paragraphs), because serifs might help guide eyes to the next line, by helping to mark vertical locations. This is a relevant issue for future research.

It is important to stress that the failure to obtain any advantage of the serif font over the sans serif font occurred despite the fact that, as readers, we are more accustomed to read text in a serif font. After all, the large majority of articles, books, e-books, etc, employ serif fonts. Importantly, this tendency is now changing. There is a current trend to employ sans serif fonts—in the same way that blackletter fonts fell into disuse in the past century. For instance, in the early versions of Microsoft Office, the default font was Times New Roman (i.e., a serif font), but in later versions, the default font has changed to Calibri (i.e., a sans serif font). Similarly, in the new era of digital technology, current e-book applications have serif fonts as a default, but many of these applications allow the user to change it to a sans serif font (e.g., iBook and other e-book applications). This movement is consistent with the already prevalent use of sans serif fonts in signage in public areas (e.g., roads, streets, buses, etc)—bear in mind that under suboptimal conditions of distance the serifs in the letters may well act as noise (i.e., under some circumstances, serifs may have a deleterious effect on letter/word recognition). For instance, in their review on fonts and legibility, Rusell-Minda et al. (2007) concluded that “with small letter sizes, close to the acuity limit, serifs may actually interfere, although slightly, with legibility” (p. 410). Indeed, the *American Printing House for the Blind* recommends the use of sans serif font for individuals with low vision: “For text, a readable typeface means a sans-serif (/san-ser-if/) typeface (or font) made up of mainly straight lines. A serif is a short stroke that projects from the ends of the main strokes that make up a character. These are not desirable for use in a book to be read by persons of all ages and/or persons with visual impairments” (http://www.aph.org/edresearch/lpguide.htm, retrieved on June 17, 2012). Similar recommendations are in use for readers with dyslexia. For instance, in the website of the British Dyslexia Association, they recommend “a plain, evenly spaced sans serif font such as Arial” (http://www.bdadyslexia.org.uk/about-dyslexia/further-information/dyslexia-style-guide.html, retrieved on June 17, 2012). Finally, as indicated in the Introduction, the fonts used in traffic signs are sans serif (e.g., *Clearview* in US roads; *Caractères* in France, *DIN 1451* in Germany, etc.)—note that these fonts were designed for increased legibility in substandard scenarios (i.e., reading traffic signs while driving).

The take-home message of the present paper is straightforward: In a normal reading setting (at least in single-sentence reading with participants with normal vision), the presence of serifs does not impact on reading fluency. The present data are entirely consistent with the claim that “reading appears to proceed at about the same rate if the type font, size, and length of line employed are at all reasonable” (Rayner, Pollatsek, Ashby, & Clifton, 2012). Thus, there is no particular reason why the APA publication norms (or any other norms) recommend the use of a serif font (Times New Roman) other than uniformity—and probably aesthetics (e.g., at a subjective level, readers may consider serif fonts as more beautifully crafted than sans serif fonts). Uniformity is the key factor here, and it is entirely reasonable that manuscripts, theses, etc., should be submitted using one/two fonts which are familiar and easily legible.

References

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Arditi, A., & Cho, J. (2005). Serifs and font legibility. *Vision Research, 45*, 2926-2933.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Blignaut, P. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics, 71*, 881-895.

Davis, C.J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117*, 713-758.

Davis, C.J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods, 37*, 665-671.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences, 9*, 335-341.

Hustwit, G. (Producer, Director) (2007). *Helvetica* [Motion picture]. United States: Plexifilm.

Merriam-Webster's manual for writers and editors (2003). Springfield, MA: Merriam-Webster

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

Morris, R.A., Aquilante, K., Bigelow, C., & Yager, D. (2002, May). *Serifs slow RSVP reading at very small sizes, but don't matter at larger sizes*. Paper presented at the SID Digest of Technical Papers symposium, Boston, MA.

Morrison, R.E., & Inhoff, A.W. (1981). Visual factors and eye movements in reading. *Visible Language, 15*, 129-146.

Perea, M., & Acha, J. (2009). Space information is important for reading. *Vision Research, 49*, 1994-2000.

Poulton, E.C. (1965). Letter differentiation and rate of comprehension in reading. *Journal of Applied Psychology, 49*, 358-362.

Rayner, K. (2009). The Thirty Fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457-1506.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. Jr. (2012). *The Psychology of Reading*. 2nd edition. New York: Psychology Press.

Reynolds, L. (1979). Legibility studies: Their relevance to present-day documentation methods. *Journal of Documentation, 35*, 307-340.

Russell-Minda, E., Jutai, J.W., Strong, J.G., Campbell, K.A., Gold, D., Pretty, L., & Wilmot, L. (2007). *Journal of Visual Impairment & Blindness, 101*, 402-415.

Sanocki, T., & Dyson, M.C. (2012). Letter processing and font information during reading: Beyond distinctiveness, where vision meets design. *Attention, Perception, & Psychophysics, 74*, 132-145. DOI 10.3758/s13414-011-0220-9.

Slattery, T.J., & Rayner, K. (2010). The influence of text legibility on eye movements during reading. *Applied Cognitive Psychology, 24*, 1129-1148.

Tinker, M.A., & Paterson, D.G. (1932). Studies of typographical factors influencing speed of reading: IX. Reduction in the size of newspaper print. *Journal of Applied Psychology, 16*(5), 525-531.