Letter and symbol identification: No evidence for letter-specific crowding mechanisms

Eric Castet  
Aix Marseille Univ, CNRS, LPC, Marseille, France

Marine Descamps  
Aix Marseille Univ, CNRS, LPC, Marseille, France

Ambre Denis-Noël  
Aix Marseille Univ, CNRS, LPL, Marseille, France

Pascale Colé  
Aix Marseille Univ, CNRS, LPC, Marseille, France

It has been proposed that letters, as opposed to symbols, trigger specialized crowding processes, boosting identification of the first and last letters of words. This hypothesis is based on evidence that single-letter accuracy as a function of within-string position has a W shape (the classic serial position function [SPF] in psycholinguistics) whereas an inverted V shape is obtained when measured with symbols. Our main goal was to test the robustness of the latter result. Our hypothesis was that any letter/symbol difference might result from short-term visual memory processes (due to the partial report [PR] procedures used in SPF studies) rather than from crowding. We therefore removed the involvement of short-term memory by precueing target-item position and compared SPFs with precueing and postcueing. Perimetric complexity was stringently matched between letters and symbols. In postcueing conditions similar to previous studies, we did not reproduce the inverted V shape for symbols: Clear-cut W shapes were observed with an overall smaller accuracy for symbols compared to letters. This letter/symbol difference was dramatically reduced in precueing conditions in keeping with our prediction. Our results are not consistent with the claim that letter strings trigger specialized crowding processes. We argue that PR procedures are not fit to isolate crowding processes.

Introduction

A peripheral target displayed with close flankers appears scrambled and is more difficult to identify than when it is displayed on its own. This phenomenon, known as visual crowding, has been extensively studied in the last two decades. Crowding is thought to be a major limiting factor of reading (Pelli et al., 2007) and of spatial vision although the precise nature of its underlying mechanisms is still strongly debated (He, Cavanagh, & Intriligator, 1996; Herzog, Sayim, Chicherov, & Manassi, 2015; Intriligator & Cavanagh, 2001; Pelli, Palomares, & Majaj, 2004; Pelli & Tillman, 2008; Saarelta, Sayim, Westheimer, & Herzog, 2009; Strasburger, Harvey, & Rentschler, 1991; Strasburger & Malania, 2013; Tyler & Likova, 2007; Wolford, 1975; Yeshurun & Rashal, 2010; Yeshurun, Rashal, & Tkacz-Domb, 2015). Some recent reviews show the difficulty of providing a clear-cut understanding of crowding (Levi, 2008; Strasburger, Rentschler, & Jüttner, 2011). Crowding has also been invoked to account for reading deficits either in amblyopia (Levi, Song, & Pelli, 2007), in low vision (Wallace, Chiu, Nandy, & Tjan, 2013; Wallace, Chung, & Tjan, 2017), or in developmental dyslexia (Martelli, Filippo, Spinelli, & Zoccolotti, 2009) although this remains a controversial issue both in low vision (Bernard, Scherlen, & Castet, 2007; Calabresè et al., 2010; Chung, 2004; Scherlen, Bernard, Calabresè, & Castet, 2008) and dyslexia (Doron, Manassi, Herzog, & Ahissar, 2015; Ramus & Ahissar, 2012).

Although many studies of crowding have investigated how the structural difference between the target and the flankers influences the strength or extent of crowding (Bernard & Chung, 2011; Chung, Levi, & Legge, 2001; Kooi, Toet, Tripathy, & Levi, 1994; Wang, He, & Legge, 2014), only a few studies have investigated whether different categories of items could induce different types of crowding (Reuther & Chakravarthi, 2014). Based on an influential study in the field of psycholinguistics, it has been claimed that letters and symbols may trigger different crowding processes (Tydgat & Grainger, 2009). This study (hereafter referred to as the TG study) has been the original departure of a line of research investigating...
whether crowding processes are modified during reading acquisition for frequent items, such as letters. Proponents of this view argue that extensive exposure to alphanumerical characters while children learn to read can create new crowding processes that are automatically triggered when strings of letters are processed. This transformation is supposed to reduce or suppress detrimental effects of crowding whenever a word or, more precisely, any string of alphanumerical characters is processed. The major evidence reported in the TG study to support these hypotheses is a dramatic difference in identification performance between letters and symbols. In the basic experiments, a five-letter string was briefly displayed followed by a mask with a postcue indicating the location of the unique letter to be reported. This is the famous partial report task used in numerous studies (Averbach & Coriell, 1961; Bundesen, Pedersen, & Larsen, 1984; Sperling, 1960, 1965). The results in the letter condition of the TG study were consistent with a wealth of previous studies showing that accuracy as a function of within-string letter position (the so-called serial position function or SPF) is a W-shape function. Note that the SPF in psycholinguistics should not be confused with the SPF used in the field of memory (e.g., Kelley, Neath, & Surprenant, 2013). This function shows that performance is optimal at fixation (middle letter) as well as at outer letters with a dramatic disadvantage for the intermediate letters (Averbach & Coriell, 1961; Butler, 1975; Butler & Merikle, 1973; Campbell & Mewhort, 1980; Merikle & Coltheart, 1972; Merikle, Coltheart, & Lowe, 1971; Merikle, Lowe, & Coltheart, 1971; Mewhort & Campbell, 1978; Mewhort, Campbell, Marchetti, & Campbell, 1981; Townsend, Taylor, & Brown, 1971).

However, the TG study also showed that the standard W shape of the accuracy function was replaced by an inverted V shape when keyboard symbols were used instead of letters. In other words, although strings of letters showed a dramatic advantage for the first and last positions (outer or end positions), this advantage was absent with symbols, i.e., outer-symbol performance was smaller than performance for the three other letters. The authors interpreted their result as evidence that crowding extent was smaller for letters (and more generally for alphanumerical characters) than for symbols. Thus they argued, “The first and last symbols in an array of symbols are subject to practically the same amount of crowding as the inner letters. Letters and digits, on the other hand, would be subject to lower levels of crowding that would depend on the number of flanking characters (one or two) because of the hypothesized narrower receptive fields of the detectors in the alphabetic array” (Tydgat & Grainger, 2009, p. 494). We will use the same terminology as the authors and refer to this model cast in terms of crowding as the “modified receptive field” hypothesis.

Before speculating on the theoretical consequences of the modified receptive field hypothesis in the fields of visual psychophysics and psycholinguistics, some important points need further investigation before concluding that crowding was responsible for the letter–symbol difference. In short, the TG study does not seem appropriate to isolate crowding processes for one crucial reason having multiple consequences. In typical crowding studies, observers know the location of the target item either before the stimulus is displayed or when the stimulus appears. For instance, observers know that the target item is always the one in the middle of a three-item string. Note that this absence of spatial uncertainty is true even in studies investigating the influence of attention on crowding (Yeshurun & Rashal, 2010). A direct consequence is that observers do not have to store in a memory buffer (see below) the whole set of displayed items: Instead, visuo-attentional resources can be focused on the location of the target item during its physical display to alleviate interference with the spatially adjacent flankers.

In contrast, in the TG study, target location is only indicated after physical disappearance of the stimulus (with spatial postcues located close to the target). This is inherent to the partial report procedure. In a partial report task, by definition, observers have to store in a memory buffer as many items as possible. It is only when the spatial postcue is displayed (i.e., after stimulus disappearance) that visuo-attentional resources can be deployed toward the target location: Note that these processes operate on the stored representation of a stimulus that is no longer physically present (Reeves & Sperling, 1986).

Thus the critical issue in the TG study is that using a partial report procedure implies short-term visual memory processes that are absent in typical crowding studies. Results based on a partial report task can therefore be interpreted with many different models having no theoretical link with crowding. Crucially, we emphasize that attention is not distributed similarly in a standard crowding task and in a partial report task. In the latter case, during the physical display of the stimulus, attentional resources are distributed in a diffuse way over all items while waiting for postcues (in addition, the center of this attentional distribution coincides with gaze fixation). It is very different from standard crowding studies in which attentional resources are spatially focused on the target location (and the center of the distribution is usually in the periphery). It is during the period of stimulus presentation that some critical differences might occur between the two experimental paradigms. For instance, some items within a category might be much more salient than the others and pop out during the stimulus
display period, thus creating a form of attentional
capture (Bundesen, 1990; Nordfang & Bundesen, 2010; 
Theeuwes, 2010). This attentional capture would thus
warp the intended diffuse attentional distribution and
induce an attentional focus at a location that would
often be away from the location of the subsequent
postcue (hence of the target). In contrast, in standard
crowding experimental paradigms, attention is effi-
ciently deployed on the target location during the
stimulus display and can thus more efficiently coun-
teract attentional capture (Lim & Sinnett, 2014),
sequential priming effects (Maljkovic & Nakayama,
1996), or idiosyncratic attentional distributions across
space. Thus, partial report procedures, compared to
standard crowding procedures, will induce complex
and uncontrolled effects related to the interaction
between attentional processes and short-term visual
memory.

Based on the numerous potential issues described
above, the general aim of the present study was to
investigate the validity of assessing crowding processes
with a partial report procedure. We wanted to compare
results obtained in the standard conditions used when
measuring the SPF (partial report) with results
obtained in similar conditions except that performance
did not rely on short-term visual memory any more.
More specifically, we wanted to test whether results
reported in the TG study would still be observed when
the involvement of short-term memory was not
necessary. This goal was achieved by precueing the
target location to remove the necessity of storing the
five items of the string in a short-term visual memory
buffer. In addition, we also tried to minimize low-level
visual differences between letters and symbols. We
hoped that increasing visual similarity of items within
each category would reduce the attentional capture
issues mentioned above. Note that attentional capture,
i.e., when attention is involuntarily directed toward an
irrelevant item (through exogenous attention), is still a
potential nuisance even when target location is precued.
For instance, a voluntary shift of attention toward the
first-position item (through precueing) might be im-
peded by an involuntary exogenous shift of attention
toward a very conspicuous item at the last position. We
addressed this problem by controlling the spatial
complexity of items to equalize as much as possible the
conspicuousness of our items. We decided to control
perimetric complexity, which is a strong determinant of
crowding processes (Pelli, Burns, Farell, & Moore-
Page, 2006). The range of spatial complexities for the
group of symbols used in the TG study was much
higher than for the group of letters (Figure 1A; see
details in the Methods section). We reasoned that this
large letter–symbol difference in within-category vari-
ability might have induced more frequent attentional
capture effects when symbols were presented. To

<table>
<thead>
<tr>
<th>Item</th>
<th>letter</th>
<th>symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/Courier New</td>
<td>S</td>
<td>$</td>
</tr>
<tr>
<td>B/Deja Vu Sans Mono</td>
<td>N</td>
<td>K</td>
</tr>
</tbody>
</table>

Figure 1. Perimetric complexity of letters and symbols. (A) Items
used in the TG study. Note the very large variability between
symbols. (B) Items used in the present study. Symbols were
created to obtain a one-to-one matching in perimetric
complexity between letters and symbols. This accurate
matching is not perfectly rendered in the figure because items
had to be displayed with a random jitter to avoid overlapping.

control for this potential nuisance, we created nine
symbols that were individually matched with nine
consonant letters based on spatial complexity (one-to-
one correspondence) so that within-category variability
of spatial complexity was identical for both groups
(Figure 1B). We hoped that this stringent control of
spatial complexity would minimize potential attentional
capture effects that could differentially affect the
two categories. This manipulation also had the
advantage of keeping the mean spatial complexity
identical for the two groups.

Two experiments were performed. In both experi-
ments, two conditions were compared in a repeated-
measures design. In the postcuing condition, short-
term visual memory was needed as target location was
indicated after the five-item stimulus disappearance:
This was the partial report procedure used in the TG
study. In the precueing condition, short-term visual
memory was not needed as target location was
indicated before the five-item stimulus appearance:
This aimed at getting closer to standard crowding
experiments while keeping the same stimuli as those
used in the TG study (i.e., five-item strings). In
Experiment 1, the two-level “cueing” factor was crossed
with a two-level intercharacter spacing factor (thus
resulting in four conditions run for all observers). The two levels of the “spacing” factor were standard spacing (as in the TG study and most studies of SPFs) and double spacing. As double spacing has been shown to degrade reading performance (Chung, 2002; L. Cohen, Dehaene, Vinckier, Jobert, & Montavont, 2008; Vinckier, Qiao, Pallier, Dehaene, & Cohen, 2011), we wondered whether these less optimal stimuli would reduce the letter-specific advantage for outer-position items found in the TG study. A second experiment was performed to test the robustness of the results obtained in Experiment 1 and to replicate more closely the temporal masking parameters used in the TG study. Experiment 2 was thus similar to Experiment 1 except that the five-item string was always preceded and followed by masks (as in the TG study) whereas only postmasks were used in Experiment 1 (intercharacter spacing was always standard in Experiment 2 as in the TG study).

To anticipate our results, using stimuli and methods very similar to those used in the TG study, the collapse in performance for outer-position symbols was not reproduced. We only observed an overall reduced performance for symbols compared to letters. This latter finding might reflect a difference in familiarity between letters and symbols, a result whose standard explanations have nothing to do with crowding (Changizi & Shimojo, 2005; Doron et al., 2015; Mewhort et al., 1981; Norris, 2013; Norris & Kinoshita, 2012; Wiley, Wilson, & Rapp, 2016; Wong, Jobard, James, James, & Gauthier, 2009). Overall, our work is not consistent with the hypothesis that letters would be processed by specialized crowding processes having the function of boosting processing of outer letters.

**Methods**

**Subjects**

Twenty young subjects with normal or corrected-to-normal vision participated in Experiment 1, and 20 different subjects participated in Experiment 2 (mean age: 22 years). All were right-handed and had French as their mother tongue. The research followed the tenets of the Declaration of Helsinki and was approved by the Ethical Committee for Protection of Human Subjects at the Aix-Marseille University (AMU). Written informed consent was obtained from each observer after the nature and purpose of the experiment had been explained. All observers were naive as to the purpose of the experiments and received a monetary compensation (20 euros for about 2.5 hours).

**Apparatus**

Stimuli were displayed on a 21-in. CRT color monitor (ViewSonic P227f, refresh rate = 85 Hz, resolution = 1600 × 1200 pixels) driven by a PC computer running custom software developed in Python with the Psychopy library (Peirce, 2007). Observers sat in a comfortable chair with their eyes at a distance of 60 cm from the monitor in a dimly lit room (screen visual angles: 37.7° × 28.3°). They viewed the screen with their head maintained by a chin rest. An eye tracker (see section Eye movement recording) was connected to our system. Letters and symbols were displayed in black (luminance: 0.3 cd/m²) on a light gray background (luminance: 60 cd/m²).

**Stimuli**

The items used in our study are shown in Figure 1 (on the right) along with the items used in the TG study (Figure 1 on the left). This figure also allows us to easily compare the perimetric complexity for all items (Pelli et al., 2006). Letters in our study were uppercase characters from the DejaVu Sans Mono font (https://dejavu-fonts.github.io/), a free, fixed-width, sans-serif font having a smaller perimetric complexity than the fixed-width, serif Courier font (Pelli et al., 2006). Each letter was drawn from a set of a nine letters (F, H, K, L, P, T, V, X, Z). Nine symbols were created with the software FontCreator (http://www.high-logic.com/font-editor/fontcreator.html). Each new symbol was designed based on one of the nine letters with the goal of keeping perimeter and area constant so that there was a one-to-one matching in perimetric complexity between the nine letters and symbols (Pelli et al., 2006). The font containing these nine symbols can be obtained from the authors on request for scientific purposes. Perimetric calculations were made with the SciKit-image package in Python (van der Walt et al., 2014). Our font has a mean perimetric complexity of 6.1 for both categories of items (SD = 1.1). In the TG study (which used the Courier New font), mean perimetric complexity was 14.5 (SD = 1.8) for letters and 13.6 (SD = 6.4) for symbols.

The height of the items was 0.76, thus corresponding to an x-height of 0.55° for the corresponding lowercase characters (Legge & Bigelow, 2011). Using a standard intercharacter spacing, this induced a center-to-center distance between adjacent characters of 0.6° (standard spacing condition). The five positions of the characters were therefore located at −1.2°, −0.6°, 0°, +0.6°, and +1.2° from the center of the screen. An additional condition was used in which the intercharacter distance was twice as large as the standard spacing while character size remained constant (re-
ferred to hereafter as the double spacing condition). In this condition, the five positions were located at $-2.4\,^\circ$, $-1.2\,^\circ$, $0\,^\circ$, $+1.2\,^\circ$, and $+2.4\,^\circ$ from the central fixation point.

**Temporal course of a trial**

A schematic view of the temporal course of a trial is shown in Figure 2 for Experiments 1 and 2.
Experiment 1 (Figure 2A)

On each trial, a small dot was first displayed in the center of the screen. Subjects had to look at this dot to make it disappear and launch the rest of the trial. This step allowed us to check that eye-tracker calibration was accurate across an entire experimental block (see the details of this step in section Eye movement recording). When the dot disappeared, two vertical lines aligned with the center of the screen were displayed. Then observers had to keep on fixating between these two lines and press a button, which triggered the display of the rest of the trial. In precueing experimental blocks, a 100% valid precueing probe (two horizontal lines above and below the target item) was displayed for 153 ms and followed by a horizontal string of five items (letters or symbols). The duration of the five-item string was 94 ms in all conditions. In postcueing experimental blocks (i.e., partial report task), the five-item horizontal string was displayed immediately after the button press. In both cueing conditions, the string of five items was followed by a backward mask of five hash signs displayed for 800 ms. In the postcueing condition, the 100% valid spatial probe (two horizontal lines above and below the target item) was displayed at the same time as the hash signs. Then, observers had to identify the target item (i.e., the item that had been cued) by clicking with the mouse on one of the nine items that were displayed on the screen. Accuracy for target item identification was the dependent variable.

Experiment 2 (Figure 2B)

In the second experiment, two masks were always used: one before the stimulus for 153 ms and one after the stimulus for 800 ms. As in Experiment 1, the spatial probe could be displayed either before the stimulus, but this time along with the premask (precueing), or after the stimulus (postcueing). The two cueing conditions were presented in different experimental blocks as in Experiment 1.

Eye movement recording

Subjects’ gaze location was monocularly recorded at a rate of 1000 Hz with an EyeLink 1000 eye tracker (Eyelink 1000 Tower Mount distributed by SR Research Ltd., Mississauga, ON, Canada). Before each session, a five-point gaze calibration was performed followed by a five-point validation. Calibration and/or validation were repeated until the validation error was smaller than 1° on average and smaller than 1.5° for the worst point. Ocular data were used online and also recorded for further verifications with the Data Viewer software (SR Research Ltd.).

Accurate online eye tracking was used to control two critical points: (a) that gaze was directed to the middle item of the string when the stimulus appeared and (b) that gaze stayed at this location during the display of the string of items.

To control for the first point, the following procedure was used. As shown in Figure 2, each trial started with the display of a fixation dot (centered on the screen). If gaze was found to be within a 1° region centered on this dot, then the rest of the trial could be displayed. Otherwise, if gaze was not present in this region for a period of 5 s, the fixation dot disappeared. This indicated either that the observer had not been fixating the dot for 5 s or that the eye tracker was not accurately calibrated any longer. At this stage (i.e., at the beginning of any trial), the operator could recalibrate the eye tracker with a new cycle of calibration/validation steps and launch the trial again.

To control for the second point, online eye tracking was used to check whether gaze stayed within a 1° region centered on the middle item during the presentation of the five-item string. If that was not the case, a sound was produced to give a warning feedback to observers. This helped them remember that keeping their eyes still during the five-item display was a requirement of the task in which they were engaged.

Design and procedure

The five-item string displayed on each trial was chosen with the following constraints. Within an experimental block, each of the nine items was a target at each of the five string locations, thus resulting in 45 trials. In any trial, for the four locations that did not contain a target, items were drawn without replacement from the nine-item set minus the target item. With the set of letters, for instance, a valid string could be F, P, Z, T, H with the letter F being the target at the first location. One experimental block thus contained 45 trials with accuracy being estimated based on nine observations for each of the five locations.

In addition to the probe-position factor, three other factors were investigated: type of character (letter vs. symbol), intercharacter spacing (standard vs. double), and cue type (precueing vs. postcueing). These three factors were blocked, and the cross-combination of their levels produced eight different experimental blocks. Two sessions, each containing these eight randomly interleaved blocks, were run with a 15- to 20-min break between the two (16 blocks overall for each subject). Eventually, each data point measuring accuracy was based on 18 observations per subject and per
factor combination (i.e., 18 \times 5 \text{ locations} \times 2 \text{ character types} \times 2 \text{ spacings} \times 2 \text{ types of cueing} = 720 \text{ repeated measures per subject}).

The two sessions were always preceded by practice blocks, run for about 10–15 min, to allow observers to become familiar with the task and with the stimuli. The whole experiment lasted approximately 2.5 hr for each of the 20 observers.

**Experiment 2**

The design in Experiment 2 was identical to that used in Experiment 1 except that the intercharacter spacing was always standard.

**Statistical analysis**

Statistical analyses were based on linear mixed-effects models specifying subjects and items as random factors (Baayen, 2008; Zuur, Ieno, & Elphick, 2010). We used the R system for statistical computing (R Core Team, 2016) with the lmer and glmer functions from the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). We also used the following additional packages for graphs, data processing, and tables, respectively: ggplot2, tidyverse, and stargazer. The Akaike information criterion and likelihood-ratio tests were used to select a parsimonious model (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017; Zuur et al., 2010). Likelihood-ratio tests were performed with the anova() function in the lme4 package. Assumptions underlying the models were visually checked with diagnostic plots of residuals (Cohen, Cohen, West, & Aiken, 2003; Pinheiro & Bates, 2000).

**Summary**

Every effort was made to ensure that important methodological characteristics were similar in the present study and in the TG study. Main similarities and differences are presented in Tables 1 and 2, respectively.

**Results**

**Experiment 1**

**Analysis of accuracy**

Responses were scored as correct or incorrect: Figure 3 summarizes the accuracy data as a function of probe position for the four conditions tested (2 cueing types \times 2 interitem spacings). In each panel of Figure 3, large squares (for symbols) and circles (for letters) connected by thick lines represent means across the 20 observers. Individual data are represented by smaller squares and circles. In the four conditions, the results clearly show the typical W-shaped serial position function for the letters as well as for the symbols. In other words, the inverted V shape reported in the TG study for symbols only is not reproduced here: The accuracy for outer symbols is systematically higher than for intermediate symbols.

Results were analyzed with generalized mixed-effects analysis. A fourth-degree polynomial regression was used to regress accuracy on probe position following
standard practice when analyzing serial position functions (Mason, 1982). As shown in Figure 4 with solid lines, using a fourth-degree regression allows us to adequately fit the W-shaped serial position function (a third-order polynomial cannot describe the W-shaped data mainly because it has at most two maxima as the first derivative is second order).

The generalized linear mixed-effects model whose predictions are shown in Figure 4 included the following fixed effects: cueing, interitem spacing, type of item, and a fourth-degree polynomial for probe position. The random effects structure contained random intercepts and cueing effects for subjects and items. Results of this model are displayed in Table 3: Note that regression coefficients are reported in logit units (log odds values with base e: \( \log \frac{p}{1-p} \)) whereas accuracy is reported in probability units in Figure 4.

All effects reported in Table 3 are statistically significant with a very small \( p \) value. In a generalized regression model, the amplitude of an effect is commonly expressed in terms of an odds ratio: here, the ratio between the odds of successful item identification for two different conditions (as a reminder, the odds of success is the ratio of the probability of success over the probability of failure). We can recover the odds ratio by exponentiating the regression coefficient provided in logit units in Table 3. Notably, the effect of postcueing compared to precueing corresponds to an odds ratio of 0.43 (i.e., \( \exp [-0.84] = 0.43 \)). This means that the odds of success in the postcueing condition are less than half the odds of success in the precueing condition; this is a very large amplitude effect. In comparison, the effect of using symbols compared to letters is smaller with an odds ratio of about 0.62 (i.e., \( \exp [-0.48] = 0.62 \)); the odds of success with symbols are about 0.6 times the odds of success with letters.

We do not comment on the individual coefficients of the fourth-order polynomial regression of accuracy on probe position because coefficients from high-order polynomial regression equations are notoriously difficult to interpret (e.g., Edwards & Parry, 1993; Mirman, 2014). The main asset of using such a polynomial is to allow a linear, rather than nonlinear, regression for a function having a relatively complex shape that is not based on theoretical grounds (Mason, 1982).

One tricky aspect of these data is the presence of conspicuous ceiling and floor effects, respectively, for positions 1 and 5 on the one hand and for positions 2 and 4 on the other hand (Figure 3). Here, floor effects...
correspond to performance near chance level, namely, around an accuracy of 0.11 (1/9), and ceiling effects correspond to the maximal accuracy of one. For instance, visual inspection of individual data at position 2 in Figure 3 shows that many observers are around chance level.

Having either floor or ceiling effects is problematic because it can induce spurious interactions. Indeed, an interaction (irrespective of its statistical significance) cannot be meaningfully interpreted when data suffer from the presence of floor or ceiling effects. This critical caveat is described in many textbooks although it is often ignored (e.g., Kantowitz, Roediger, & Elmes, 2014). And this issue is also notoriously compounded with noncrossover interactions, i.e., similar to those observed here (Loftus, 1978; Wagenmakers, Krypotos, Criss, & Iverson, 2012). For these reasons, the statistical models in the present work do not include interactions.

Another noticeable point about the precueing data presented in Figure 3 is the difference in accuracy between the central (and fixated) probe position (p3) and the outer positions (p1 and p5). The smaller accuracy at p3 compared to p1 and p5 is not intuitive both because p3 is the position of highest accuracy and because it is precued. This pattern could be accounted for by two nonexclusive factors. First, foveal crowding or contour suppression (Lev, Yehezkel, & Polat, 2014) might be relatively stronger compared to outer positions if only because p3 has two “direct” flankers whereas p1 and p5 only have one “direct” flanker. More precisely, p3 has four flankers, two to the left and two to the right, a situation inducing more crowding than only two direct flankers (Strasburger et al., 1991). Second, there is clear evidence that the detrimental effect of a backward mask (having the same size as the string of items) on the SPF is more pronounced at the central position than at the outer positions (Mewhort & Campbell, 1978; Mewhort et al., 1981).

**Analysis of errors**

As our analysis of accuracy results showed a dramatic effect of precueing compared to postcueing, we compared the different patterns of errors obtained in the two conditions to better characterize the effects of attention. This kind of error analysis was introduced about 50 years ago (Eriksen & Rohrbaugh, 1970)—see also Wolford (1975)—and introduced in the crowding literature by Strasburger et al. (1991). Following previous work on the SPF, errors were classified as either intrusions or location errors (Mewhort et al., 1981). An intrusion error occurs when the reported item is not contained within the displayed string of five items.
items. A location error occurs when the reported item is one of the four adjacent items in the displayed string.

In the crowding literature, location errors are similar to localization errors or whole-letter confusions (Chung & Legge, 2009; Huckauf & Heller, 2002; Levi, 2008; Strasburger, 2005, 2014; Mewhort & Malania, 2013). In addition, a new result is that the main effect of precueing is to decrease intrusion errors. This is a large amplitude effect. A generalized linear mixed-effects analysis shows that the odds ratio of failure in the postcueing condition, compared to the precueing condition, is very large (estimate: 1.01 logits, SE: 0.06, p < 0.0001). This means that the odds of making intrusion errors are multiplied by 2.75 in the postcueing condition. In comparison, for location errors, the odds of failure are only multiplied by 1.25 in the postcueing condition.

Another conspicuous result emerges from the analysis of location errors as shown in Figure 6. For each probe position, the probabilities of reporting an item displayed at one of the four other positions within the string is reported. The most striking result is observed for the second and fourth probe positions and is especially conspicuous in the precueing condition. When the probe position is at the second position, i.e., when the target item is at the second position, the most frequent location error consists in reporting the letter displayed at the first position. Similarly, when the probe position is at the fourth position, the most frequent location error consists in reporting the letter displayed at the fifth position.

### Experiment 2

To assess the robustness of the results presented in Experiment 1, we felt that it was important to replicate our own results with a new set of 20 subjects. We also wanted to check whether the pattern of our results might have been influenced by the masking characteristics of our stimuli in Experiment 1: Whereas premasks and postmasks were jointly used in the TG study, only postmasks were used in our first experiment. We therefore decided to use premasks and postmasks in Experiment 2 (see Figure 2B).

Results are shown in Figure 7 with the same characteristics as those used to report the results of Experiment 1 in Figure 3. The pattern of results is strikingly similar to that observed in Experiment 1 (results of the mixed-effects analysis are shown in Table 4). First, the W shape is again clear-cut for symbols in the two cueing conditions. Second, the difference between letters and symbols is very small or absent, depending on positions in the precueing condition.

### Discussion

Based on studies using partial report tasks, some authors have claimed that symbols and letters do not trigger the same crowding processes although the

<table>
<thead>
<tr>
<th>Dependent variable: Accuracy in logit units (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (precueing, 1 × spacing, letters, position 3)</td>
</tr>
<tr>
<td>Postcueing</td>
</tr>
<tr>
<td>2 × spacing</td>
</tr>
<tr>
<td>Symbols</td>
</tr>
<tr>
<td>Linear term for probe position</td>
</tr>
<tr>
<td>Quadratic term for probe position</td>
</tr>
<tr>
<td>Cubic term for probe position</td>
</tr>
<tr>
<td>Quartic term for probe position</td>
</tr>
</tbody>
</table>

Table 3. Experiment 1: Results of the generalized mixed-effects analysis for the fixed effects using a fourth-degree polynomial regression of accuracy on probe position. Notes: The probe-position factor was mean-centered. Estimates of regression coefficients, in logit units, are followed on the same line by standard errors in parentheses; corresponding t values are shown on the respective lines below. Accuracy is recovered by transforming back the logit estimates to probabilities with the inverse logit function (i.e., the logistic function). For instance, the intercept logit value of 0.75 corresponds to a probability of \( p = 0.68 \). * \( p < 1e-04; ** \( p < 1e-05; *** \( p < 1e-06.\)
important question of how low-level visual cortical areas are able to decide whether an item should be processed by symbol- or letter-specific crowding mechanisms was left as an open question. The initial result underlying this hypothesis was a symbol-specific collapse of accuracy at the outer positions of a five-item string as reported in the TG study (Tydgat & Grainger, 2009). In contrast, letters were shown to have the standard W shape reported in many studies, i.e., performance for the outer letters was much better than for intermediate letters (Averbach & Coriell, 1961; Butler, 1975; Butler & Merikle, 1973; Campbell & Mewhort, 1980; Merikle & Coltheart, 1972; Merikle, Coltheart, et al., 1971; Merikle, Lowe et al., 1971; Mewhort & Campbell, 1978; Mewhort et al., 1981; Townsend et al., 1971). Note that the term “crowding” was not used in these initial studies. Therefore, the “crowding hypothesis” in the TG study introduced a new terminology and framework to account for SPFs. In addition, the authors called their crowding hypothesis the modified receptive field hypothesis, thus emphasizing the claim that letters are not processed by “normal” crowding fields. Instead, letters were supposed to be processed by modified fields that have evolved during reading acquisition (usually during childhood). By this logic, only symbols or any non-alphanumeric characters would trigger normal crowding processes.

Our initial goal was to question this interpretation cast in terms of crowding. Our rationale was that the partial report procedure used in the TG study as well as in many studies from the same group may not be an adequate way of isolating crowding mechanisms for several reasons. A partial report procedure aims at investigating a form of short-term visual memory known as iconic memory or informational persistence (Coltheart, 1980a, 1983; Gegenfurtner & Sperling, 1993). Iconic memory must be distinguished from visible persistence although these two kinds of visual buffers can coexist (Castet, 1994; Castet, Lorenceau, & Bonnet, 1993; Coltheart, 1980b). In addition, theories of visual short-term memory usually distinguish iconic memory with its high-capacity buffer lasting about 300–500 ms and working memory, a low-capacity store lasting seconds to minutes (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). One still ongoing key issue for vision scientists is to understand what factors affect both the buildup and decay of iconic memory (e.g.,
For instance, it is known that modifying attentional load during the stimulus display disrupts iconic memory formation and/or decay and thus affects performance in a partial report task (Bradley & Pearson, 2012; Persuh, Genzer, & Melara, 2012). In short, a partial report procedure aims at revealing heterogeneous processes occurring after stimulus offset and corresponding to “informational persistence” (Averbach & Coriell, 1961; Bundesen et al., 1984; Coltheart, 1980a, 1980b, 1983; Sperling, 1960, 1965). Thus, any letter–symbol difference observed in the standard SPF might simply indicate that symbols and letters do not behave similarly in terms of information persistence at any stage of the processing hierarchy. In addition, in the partial report procedure, spatial attention is deployed in a diffuse way when the stimulus is displayed (i.e., before the cue indicating target location is displayed) whereas spatial attention is focused on target location in standard crowding experiments (some authors have emphasized the links between diffuse attention and ambient vs. focal processing; Previc, 1998). Some of the inadvertent consequences of these attentional differences and of their interaction with short-term visual memory when interpreting SPFs in terms of crowding have been described in the Introduction and have guided the design of our experiments.

We therefore decided to reinvestigate the TG study by controlling and investigating several factors that might have caused the dramatic letter–symbol difference, i.e., the inverse V shape for symbols. In our two experiments, perimetric complexity was controlled so that our sets of letters and symbols had the same mean value and variability (Figure 1B). Using these stimuli, the effect of spatially selective attention was investigated in the two experiments: We initially wanted to test whether precueing target location would reduce or suppress this large letter–symbol difference. In both experiments, we did find that precueing effectively reduced the difference between letters and symbols. However, the main finding was that our experiments could not reproduce the initial result of the TG study: With our symbols, there was no collapse of accuracy at the outer positions. This lack of reproducibility seems difficult to explain in terms of the stimuli being too different between studies (Table 1). The most important low-level visual characteristics of our stimuli (notably angular values of character size and intercharacter
spacing) were very similar, at least in terms of their nominal values, to those initially used in the TG study. One exception was that polarity in the TG study was based on white letters on a black background. Another issue might come from the absence of a chin rest in the TG study (J. Grainger, personal comm., October 2016). This might have induced large differences between the nominal and actual angular values in the TG study as some observers might have chosen to place their head close to the screen while other observers might have preferred to sit in their chair very far from the screen. It is possible that the very small angular values obtained when sitting away from the screen might have induced symbol-specific effects. The methods we used to display the stimuli and the experimental designs were also very similar to those used in the TG study. It must be emphasized that the inverse V curve for symbols was observed in the TG study with a two-alternative, forced-choice task as well as with an identification task (see their experiments 5 and 6) although these two tasks relied on a partial report procedure in both cases in the sense that target location was always indicated after stimulus display. There is perhaps a relevant difference that bears on the way in which observers gave their responses. In the identification task of the TG study, subjects had to look at the keyboard after each trial and press a key to report which keyboard symbol they had identified. In our work, observers gave their identification response by clicking with a mouse on one of the items directly displayed on the screen after each trial.

Table 4. Experiment 2: Results of the generalized mixed-effects analysis for the fixed effects using a fourth-degree polynomial regression of accuracy on probe position. Notes: The probe-position factor was mean-centered. Estimates of regression coefficients, in log units, are followed on the same line by standard errors in parentheses; corresponding t values are shown on the respective lines below. *p < 1e-04; **p < 1e-05; ***p < 1e-06.

<table>
<thead>
<tr>
<th></th>
<th>Intercept (precueing, letters, position 3)</th>
<th>Postcueing</th>
<th>Symbols</th>
<th>Linear term for probe position</th>
<th>Quadratic term for probe position</th>
<th>Cubic term for probe position</th>
<th>Quartic term for probe position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.68* (0.17)</td>
<td>-0.88*** (0.12)</td>
<td>-0.28*** (0.05)</td>
<td>5.45 (2.34)</td>
<td>60.48*** (2.35)</td>
<td>-10.44** (2.30)</td>
<td>37.12*** (2.22)</td>
</tr>
<tr>
<td></td>
<td>t = 3.98</td>
<td>t = -7.13</td>
<td>t = -5.25</td>
<td>t = 2.33</td>
<td>t = 25.69</td>
<td>t = -4.54</td>
<td>t = 16.74</td>
</tr>
<tr>
<td>Observations</td>
<td>7,200</td>
<td>4,202.24</td>
<td>8,430.47</td>
<td>8,519.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-272.14</td>
<td>-292.24</td>
<td>-240.97</td>
<td>-231.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akaike inf. crit.</td>
<td>8,403.47</td>
<td>8,403.47</td>
<td>8,403.47</td>
<td>8,403.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian inf. crit.</td>
<td>8,519.94</td>
<td>8,519.94</td>
<td>8,519.94</td>
<td>8,519.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7. Results of Experiment 2. Characteristics of this figure are the same as those in Figure 3.
Although this methodological difference might have some consequences, it is difficult to explain how it might have induced the inverse V shape for symbols in the TG study. Another potentially relevant difference concerns our rigorous control of gaze location. We checked before each trial whether eye-tracker calibration was correct and performed another calibration if needed. Most importantly, observers received an online feedback (an auditory signal) if their gaze moved away from fixation during the five-item display. This probably induced a much more controlled attentional deployment than in previous studies, especially when considering the tight link between attention and saccades (Castet, Jeanjean, Montagnini, Laugier, & Masson, 2006; Filali-Sadouk, Castet, Olivier, & Zenon, 2010; Montagnini & Castet, 2007). Finally, Experiment 2 showed that our results are robust and do not depend on the absence or presence of a premask (the latter case corresponding to the condition used in the TG study).

It is important to note that the precueing (vs. postcueing) manipulation was very effective in our work. This is best seen when considering the errors made by subjects (Figures 5 and 6). The general pattern is that precueing the target location dramatically decreases and actually nearly removes intrusion errors. This supports the view that location errors are the main determinants of performance in the precueing condition. In addition, the pattern of these location errors is quantitatively different in the precueing condition compared to the postcueing condition: There is a very strong tendency with precueing to report outer-position items when target location is at the intermediate positions. This result, mainly observed in our precueing condition, seems compatible either with the smaller number of flankers at the outer positions (thus inducing less crowding) or with the inward–outward visibility asymmetries observed in classic crowding studies (Bouma, 1970; Korte, 1923; Krumhansl & Thomas, 1977; Legge, Mansfield, & Chung, 2001; Petrov & Meleshkevich, 2011a, 2011b). As these signatures of crowding are much more predominant in our precueing condition, this suggests that the precueing condition is more appropriate than the postcueing condition to investigate standard crowding mechanisms. The location errors reported here might seem inconsistent with recent evidence that confusions with the inward, but not the outward, flanker increase linearly with eccentricity (Strasburger & Malania, 2013). However, these recent results were obtained with trigrams displayed at different eccentricities (the target was the middle letter), implying that the inward item was never flanked by a more central item. This is different from the visual display used in our work and does not allow an easy comparison.

Of course, as in standard crowding studies, it still remains unclear whether location errors reflect location encoding issues or response biases (Ester, Klee, & Awh, 2014; Strasburger, 2005). Another open question for future research is whether (and to what extent) the crowding mechanisms responsible for the standard inward–outward asymmetry are also responsible for the better performance at outer positions observed in the SPF (i.e., with postcueing). At first sight, this seems inconsistent with a study showing that the inward–outward asymmetry disappears when attention is allocated in a diffused way (Petrov & Meleshkevich, 2011b). There are, however, many methodological differences between this study and SPF studies (beyond the fact that this study does not rely on short-term visual memory) that could explain this apparent inconsistency. One important difference is the number of flankers, a factor known to modulate crowding strength (Strasburger et al., 1991). In SPF studies, in contrast to standard crowding studies, there is a strong asymmetry in the number of flankers around each of the target positions. This asymmetry in the relative numbers of flankers for intermediate and outer items might thus favor the outer items even in diffuse attention (postcueing) conditions. All these differences should be controlled for to investigate the links between the inward–outward asymmetry observed in standard crowding and the outer–inner asymmetry observed in SPFs.

Overall, the clear-cut differences between our precueing and postcueing conditions results emphasize that using the partial report task does not allow researchers to measure (isolate) the same processes as those that are investigated in standard crowding experiments (Herzog et al., 2015; Levi, 2008; Pelli & Tillman, 2008; Strasburger et al., 2011).

In conclusion, when we initially designed our experiments, we aimed at reducing the symbol-specific collapse of performance at the outer positions reported in the TG study (i.e., the inverted V shape). However, this collapse was not reproduced in our study whatever the conditions and experiments. SPFs in our work clearly show that items at outer positions, whether letters or symbols, have a level of performance that is much higher than at intermediate positions. It seems difficult to reconcile this basic finding with the notion that outer letters would be processed by specialized crowding processes that would boost their processing and outer symbols would be dramatically disadvantaged because they would be processed by “normal” crowding processes. The present results are consistent with the more parsimonious and traditional view that symbols are just less optimal stimuli for the visual system than letters presumably because of differences in structural properties or in expertise (Changizi & Shimojo, 2005; Doron et al., 2015; Krueger, 1975; Mewhort et al., 1981; Norris, 2013; Norris & Kinoshita, 2012; Wiley et al., 2016; Wong et al., 2009). The
present work also shows that using the partial report task may not be suitable for isolating crowding processes per se. This caveat is all the more important to bear in mind as partial report tasks are often used along with temporal masks in psycholinguistics. Unfortunately, temporal masks are notorious to induce complex spatiotemporal effects that interact with the stimulus either by adding new forms of crowding (Lev et al., 2014; Vickery, Shim, Chakravarthi, Jiang, & Luedeman, 2009) or by disrupting location encoding of items within the string (Campbell & Mewhort, 1980; Mewhort & Campbell, 1978; Mewhort et al., 1981).

Keywords: crowding, partial report, short-term visual memory, letter, symbol

Acknowledgments

We thank Thierry Ripoll for helpful discussions about previous versions of the manuscript. We further thank Hans Strasburger and an anonymous reviewer for their thorough constructive comments. Ambre Denis-Noël was supported by a PhD grant from the Brain & Language Research Institute (BLRI).

Commercial relationships: none.
Corresponding author: Eric Castet.
Email: eric.castet@univ-amu.fr.
Address: Aix Marseille Univ, CNRS, LPC, Marseille, France.

References


increased vertical word spacing in normal peripheral vision. *Optometry and Vision Science, 81*, 525–535.


foveal crowding? Scientific Reports, 4, 4067, doi:10.1038/srep04067.


Petrov, Y., & Meleshkevich, O. (2011b). Locus of


Wallace, J. M., Chiu, M. K., Nandy, A. S., & Tjan, B.


