Vision Research 244 (2026) 108822

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The influence of similarity, sensitivity and bias on letter identification

Hatem Barhoom [a,b,*,1], Mark A. Georgeson [b,c,1], Mahesh R. Joshi [b], Gunnar Schmidtmann [b]

[a] Centre for Visual Science, University of Rochester Medical Centre, Rochester, NY, USA
[b] Eye & Vision Research Group, School of Health Professions, University of Plymouth, Plymouth, UK
[c] School of Health & Life Sciences, Aston University, Birmingham, UK

ARTICLE INFO

Keywords:
Bias
Similarity
Sensitivity
Noisy template model
Letter identification

ABSTRACT

Previous studies have demonstrated that bias, sensitivity and similarity between letters are causes of errors in letter identification. However, these factors and their relative contribution in letter identification have not been investigated extensively. Our previous model (noisy template model) was devised to calculate the effect of bias and sensitivity on letter identification. In the current study, we used the method of constant stimuli to assess letter identification and the pattern of errors for Sloan letters with a range of sizes at an eccentricity of 7 deg from fixation (temporal visual field). Similar to our previous work, we devised and tested a variety of models to estimate the joint role of bias and sensitivity but extended our model to also incorporate the similarity between letters. The Modelling results revealed that bias is the major factor in determining the pattern of total, correct and incorrect responses in letter identification. Furthermore, the joint effect of similarity and bias was found to be higher than the joint effect of either bias and sensitivity or similarity and sensitivity in shaping the pattern of overall responses in letter identification. Incorporating the similarity factor into the noisy template model improved our understanding of the *simultaneous* contribution of bias, sensitivity and similarity between letters in letter identification.

1. Introduction

By their nature, letters have a complex structure and many studies have shown that legibility is different for different letters (Sloan, 1959; Bouma, 1971; Ferris et al.,1982; Grimm et al., 1994; Alexander et al., 1997; Reich & Bedell, 2000; Shah et al., 2012; Hamm et al., 2018; Ludvigh, 1941; Strasburger et al., 2011; Hairol et al., 2015; Anderson and Thibos, 2004; Shah et al., 2011). Factors affecting legibility of a given letter include its perceivability, its similarity with other letters, and the bias towards or against it (Mueller & Weidemann, 2012). Perceivability is a measure of how legible the letter is depending solely on the characteristics of the letter, such as its size, contrast, or shape. To avoid ambiguity regarding the term perceivability, we shall use the term *sensitivity* instead, which is more appropriate in light of the signal detection theory framework employed here. Response *bias* is defined as the tendency to favor one response over the other alternatives (Macmillan & Creelman, 1990), and *similarity* is defined as the confusion in letter recognition which arises among certain letters (e.g., C and O, or N and H). Letter identification could thus be affected by change in sensory input, e.g., size and contrast (sensitivity), the bias towards certain letters in case of uncertainty (response biases), and the confusion between similar letters. Note that from these definitions it is well understood that response biases, unlike sensitivity and letter similarities, are independent of the stimulus.

Luce's choice model (1963) has been frequently employed to disentangle the roles of similarity and bias on letter identification (Townsend, 1971 a & b; Gilmore et al., 1979; Mueller & Weidemann, 2012). The procedure starts by collecting response data in a letter identification task at the threshold level of test letter size. The data are collated as a stimulus-response confusion matrix (CM), where each cell represents the number of times a given letter was chosen in response to the letter presented. The model is used to decompose the CM into a bias vector and letter similarity matrix. The bias vector contains the bias parameter for each letter used as a response alternative in the experiment where the average of all bias parameter values results in the unbiased guessing rate of the task. In other words, the response bias for a given letter is the difference between observed letter usage and unbiased guessing rate (Luce, 1963; Mueller & Weidemann, 2012). The similarity matrix shows the similarity parameters for each letter pair assuming that the similarity is symmetrical (i.e., if H is similar to K, then K is similar to H by the same

---

[*] Corresponding author at: Centre for Visual Science, University of Rochester Medical Centre, Rochester, NY, USA.
E-mail address: hbarhoom@ur.rochester.edu (H. Barhoom).
[1] These authors contributed equally to this work (co-first authors).

https://doi.org/10.1016/j.visres.2026.108822
Received 30 October 2025; Received in revised form 22 March 2026; Accepted 24 March 2026

amount). However, Luce's choice model does not account for the sensitivity as it is assumed that the sensitivity of a given letter is the average of Luce's similarity parameters of each letter with the other letters (Mueller & Weidemann, 2012).

Pelli et al. (2006) investigated letter identification efficiency for a wide range of observer ages and under various conditions. They concluded that consistent letter identification efficiency across a wide range of conditions suggests that the letter identification task is a fundamental visual process (i.e., not a language skill). To identify a letter, the process starts with identifying the *features* of the letters as opposed to whole-letter templates (Pelli et al., 2006). The letter features are detected by a set of feature detectors (i.e., channels) in the visual cortex (Graham, 1989). These features are considered as crucial factors that distinguish letters from each other and therefore are key factors in letter identification (Townsend, 1971 a & b; Gilmore et al., 1979; Fiset et al., 2008). Based on the same assumption, the similarity between two letters can be estimated by determining the common features of the two letters. Here, we suppose that in principle similarity between two letters is determined by the similarity in their set of features (Pelli et al., 2006; Fülep et al., 2017). But since there is no general agreement on what those features are, or how they should be quantified, we supposed, like other investigators, that letters sharing common features (e.g., C and O; similar letters) would also show a higher *pixel-wise* correlation than letters with fewer common features (e.g., V and D; dissimilar letters).

It is important to incorporate the similarity between letters in letter identification models. The challenge is how to incorporate the three factors (bias, sensitivity and similarity) simultaneously in a single model. Mueller and Weidemann (2012) introduced a model to estimate the role of similarity in addition to sensitivity (referred to as perceivability) and response bias simultaneously. They found that all three factors were important in the letter identification task and a model with all three parameters performed better than the two-factor models that only use bias/sensitivity or bias/similarity (Mueller & Weidemann, 2012). However, their model was devised for a two-alternative forced choice (2 AFC) paradigm for letter identification at threshold level. The 2 AFC differs from the letter naming procedure in that it tends to reduce and, hence potentially underestimate bias (Macmillan & Creelman, 2005). Furthermore, the 2 AFC paradigm is influenced only by the similarity between the two presented alternatives, as opposed to the letter naming procedure (i.e., single-interval 10 AFC, employed here), where similarity can be estimated between the presented and all the other alternatives (e.g., 10 letters). Additionally, letter naming procedures are more intuitive and frequently encountered in daily life and optometric assessments.

Previously, we investigated the role of response bias and sensitivity in letter identification tasks by deriving a model (Noisy Template Model, NTM) based on the framework of signal detection theory in a single-interval 10 AFC letter identification task (Georgeson et al., 2023). The main results suggested that response bias towards or against letters and sensitivity differences between letters both influence individual letter identification. Additionally, response bias was the key factor in determining *letter usage* – how often individual letters were used as responses in the experiment, no matter whether correctly or incorrectly. However, Georgeson et al. (2023) did not estimate the effect of similarity between letters on letter identification. The current study is a continuation of this work where we introduce a novel and significant extension to the model which allows us to reveal the joint effect of bias, sensitivity and similarity in a single-interval 10 AFC letter identification task.

2. Methods

2.1. Participants

Data were collected from 12 observers (six females, mean age 34.25 ± 6.43 (SD), age range: 27–45 years) with healthy eyes and normal or corrected to normal visual acuity. Written informed consent was obtained from all observers. The study was approved by the University of Plymouth Ethics Committee. The study was conducted in accordance with the Declaration of Helsinki.

2.2. Stimuli

The stimuli used in this study were similar to the ones used in our previous work (Barhoom et al., 2021; Georgeson et al., 2023). We used black letters (luminance = 2.2 cd/m^2) on a white background (luminance = 215 cd/m^2), resulting in 99% Weber contrast. A total of 10 standard Sloan letters (C, D, H, K, N, O, R, S, V, Z) were employed. The structure of the letters follows the Sloan letter design so that their height is equal to their width and five times the stroke width.

2.3. Apparatus

The stimuli were presented on a laptop monitor (MSI, GS76 Stealth) with a resolution of 2560 × 1440 and a refresh rate of 240 Hz. Observers viewed the targets at a viewing distance of 70 cm, while sitting on a chair without using a chin or forehead rest. The examiner monitored the viewing distance by regular checks. At this viewing distance, one pixel subtended a visual angle of 0.73 min of arc ($'$). Experiments were carried out under a room illumination of ∼ 160 lx. The observer responded by calling out the responses. The responses were entered by the experimenter via a standard computer keyboard to minimize errors caused by mistyping and to improve fixation compliance. Fixation compliance was observed by the experimenter who sat opposite the observer.

2.4. Software

MATLAB 2024a (The MathWorks, Inc., Natick, Massachusetts, USA) was used to generate the stimuli, implement model fitting and statistical analysis. Routines from the Psychtoolbox-3 library were used to generate and present the stimuli (Brainard, 1997; Pelli & Vision, 1997; Kleiner et al., 2007).

2.5. Procedure

2.5.1. Data collection

The experiment was conducted monocularly. The eye was chosen at random. The fellow eye was occluded using an eye patch. The method of constant stimuli was used to collect the data. The letter stimuli were presented for 250 ms and accompanied by an auditory signal. During the experiment, the observers were asked to fixate on a fixation cross (dimensions: length and width = $11.68'$, stroke width = $1.46'$) presented at the center of the screen (Fig. 1) The letters were presented at a paracentral location with an eccentricity of 7 deg from fixation (temporal visual field). All observers were corrected for their foveal refractive error only, as the difference in optical correction between the fovea and 7 deg eccentricity is negligible (Navarro, 2009). Since letter size varied, a peripheral location was chosen (instead of a central location) to avoid presenting letters that were too small, potentially exceeding the resolution limit of the monitor. Multiple pilot experiments were conducted to establish appropriate letter sizes to cover the whole range of responses (ranging from guessing (10%) to certain decisions (100%)). Five different letter sizes (always defined by their stroke width in minutes of arc, and spaced logarithmically) were tested; $1.50'$, $2.20'$, $3.24'$, $4.76'$, and $7'$. At each size, each letter was presented 10 times. The Sloan letters with different identities and sizes were presented in random order in an interleaved design. Each observer completed 500 trials (10 Sloan letters × five letter sizes × 10 presentations per letter). The response data were collated as the number of times each possible letter was given as a response to the presented letter. Consequently, for each observer, a CM of the presented *vs* the responded letters was obtained for each letter size. Only choices of the 10 Sloan letters were accepted. To familiarize the observers with the Sloan letters, the experimenter demonstrated the

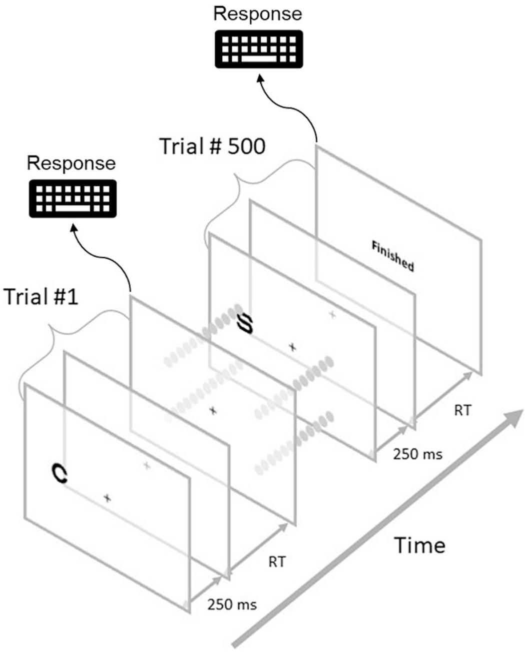

Fig. 1. Shows the sequence of events in an experimental trial (RT: response time).

Sloan letters at the beginning of the session. In the rare cases where observers responded with other letters, the experimenter reminded the observer by presenting the randomized Sloan letters set on the screen. The experimenter prompted a second response from the presented Sloan letter set. After the observer made the choice, the experimenter ensured that the observer had re-fixated before resuming the experiment. Observers showed excellent compliance in responding from the Sloan letter set (on average less than six errors per observer in 500 trials).

3. The noisy template model (NTM)

3.1. Outline of the model

The model presented in Fig. 2 is an extended version of our previous NTM (Georgeson et al., 2023). Previously, we made two simplifying assumptions about the templates. Firstly, there are as many letter templates as there are test letters in the experiment (i.e., 10 templates). Secondly, the templates were assumed to give a response only to their own preferred letter, with no response to other letters (Fig. 3A); that is, the templates were orthogonal. However, in the current version of the model, the second assumption of the NTM (i.e., the orthogonality assumption) has been revised as follows: to account for similarity between the letters, the templates respond not only to their preferred letter but also to other, similar letters. That is, the templates are correlated (i.e., non-orthogonal). When a given letter is presented, the output of each letter's template (i.e., template responses in Fig. 2) is subject to bias, sensitivity, and/or similarity to the presented letter. Furthermore, the net output of each template is perturbed by additive Gaussian noise. The bias affects the mean output level of the template by a constant amount whether the template's preferred letter is presented or not (Fig. 3B). The sensitivity affects the mean output level of the template when the preferred letter is presented (Fig. 3C). The similarity influences the mean output of the non-preferred letter templates according to their correlation to the preferred letter template (i.e., similar letters) (Fig. 3D).

3.2. Model structure and equations

For bias and sensitivity modelling, the structure and equations of the model remained the same as in the original NTM (Georgeson et al., 2023). The current experiment showed a consistent letter usage pattern similar to our previous work (Fig. 4). If differences in letter usages are

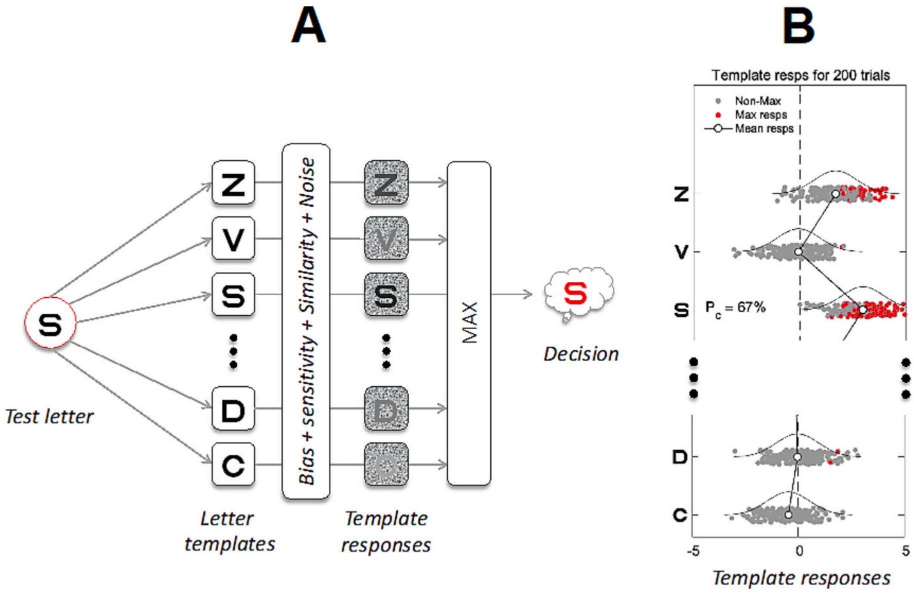

Fig. 2. A) a schematic representation of the NTM and (B) template responses. After incorporating the bias, sensitivity, similarity, and noise into the corresponding templates, the most active template (giving the MAX response) on a given trial determines the letter choice. For illustration purposes, the red points represent those trials on which a given template gave the maximum response, whereas the grey points are activations that were lower, hence not chosen. Because of noise, the most active template over trials (red dots in B) may be the correct (e.g., S) or an incorrect letter (e.g., Z). The joint effect of the three factors, bias, sensitivity and similarity determines the mean response of each template. For example, positive bias (top three rows) gives a higher mean response compared to negative bias (bottom two rows), hence, increasing the chance of incorrect responses (e.g., red dots for Z) and correct responses (i.e., S). Note that because of the dissimilarity between letter V and S, the mean response of letter V was low (although it has positive bias), therefore, it is unlikely for the template V to be chosen (as an incorrect response in this case). Negative biases (bottom two rows) decrease the chance of these letters being called, correctly or not. However, the similarity between the letters S and D increases the likelihood of letter D (although it has a negative bias) being chosen (as an incorrect response in this case).

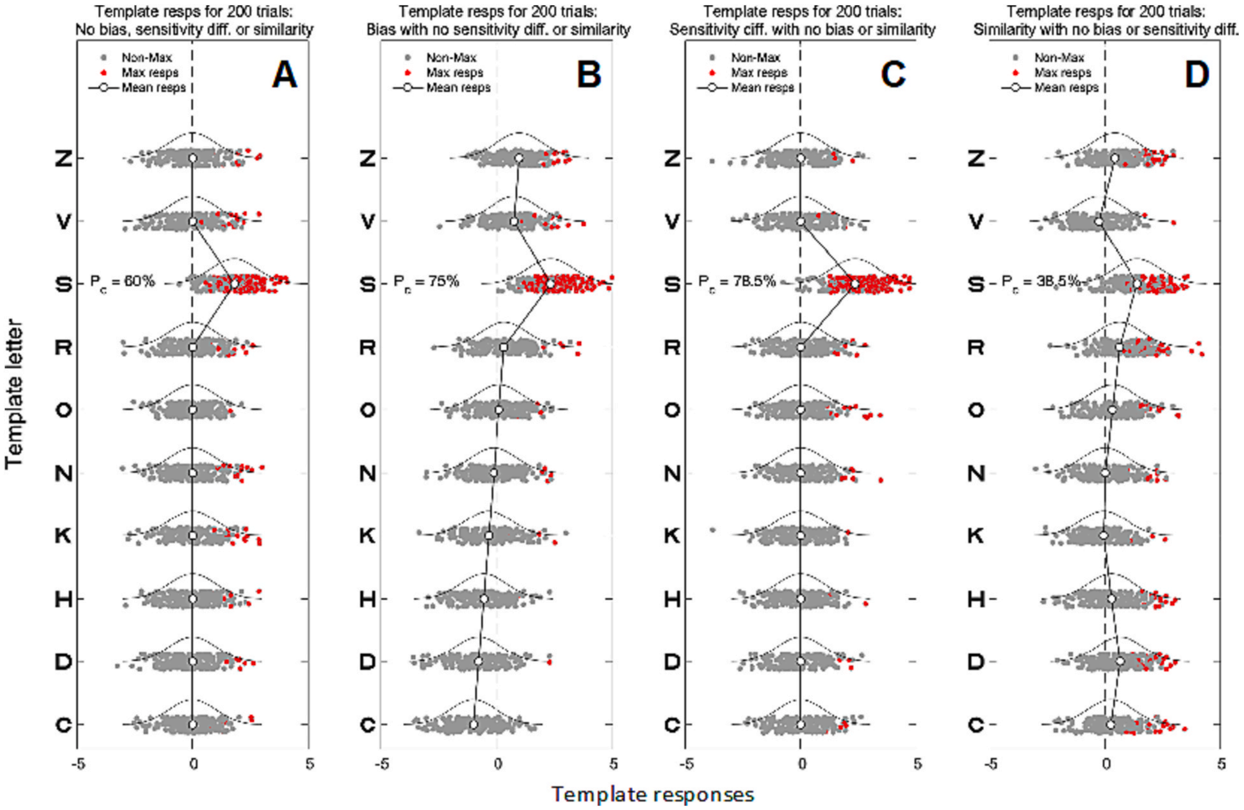

Fig. 3. Demonstration of NTM illustrating the effects of letter bias, sensitivity difference, similarity, or none of those effects. (A) the NTM without bias, sensitivity differences, or similarity between letters. (B) with different biases on each template. (C) with different sensitivity on each template. (D) with similarity between templates. The bias and sensitivity in panels B & C are ordered from negative to positive with a mean of zero. We call these the bias gradient B' and the sensitivity gradient S' (see text). The letter identity for each template here is arbitrary, for illustration only. Presenting a test letter (e.g., S) increases the mean response for the S-template but leaves others unchanged (A). The letter decision on a given trial is made by choosing the template with the largest response on that trial (the MAX operator, Fig. 2A). Choices vary from trial to trial because of noise in each template channel (indicated by the Gaussian distribution). Notice how the positive bias (e.g., Z), and presentation of the preferred letter (e.g., S) increase the likelihood of choosing certain letters, sometimes correctly, sometimes not (B). In case of sensitivity differences (C), higher sensitivity increases the chance of the letter template being chosen when its preferred letter is presented (as shown by the increase of percentage correct (P_C) from 60% to 78.5%). Panel D shows the case with no bias or sensitivity differences but incorporating similarity between the presented letter S and the other templates. This similarity decreased the likelihood of correctly choosing the letter S (P_C decreased from 60% to 38.5%) and increased the likelihood of choosing similar letters such as D or Z (i.e., incorrect responses). On the other hand, the dissimilarity between S and V decreased the likelihood of choosing V (i.e., decreased incorrect responding).

caused by biases, then the most used letter would be associated with the highest positive bias and vice versa. Suppose that the templates can be rank-ordered from least- (most-negative) bias to most-positive bias, indexed by $i = 1$ to m (where $m = 10$). Furthermore, bias values B_i (assigned to templates $i = 1$ to m) are assumed to be a linear function of i, ranging from $B_1 = -B'$ to $B_m = B'$. In this case, the number of free parameters per observer becomes one (i.e., bias parameter (B')) instead of nine.

Therefore, the bias B_i for the i^{th} template is

$$B_i = \frac{B'(2i - m - 1)}{m - 1}. \tag{1}$$

where B' is a free parameter. Subsequently, the bias values B_i are assigned to the corresponding templates according to the rank order of their letter usages in the experiment. Therefore, the highest bias value is assigned to the template of the most used letter and the lowest bias to the least used. The same procedure is followed to assign the rest of the bias values to the corresponding templates. Note that when $B' = 0$, the system is unbiased.

Similarly, we assumed that sensitivity differs (a linear variation) between templates around a baseline value (i.e., overall sensitivity S_0 which is a free parameter) ranging from $S_0 \times (1 - S')$ to $S_0 \times (1 + S')$ where S' is the sensitivity gradient. Therefore, the sensitivity of the i^{th} template to the j^{th} test letter is described by Eq. (2) and (3) for the cases where $i = j$ and $i \neq j$ respectively:

$$S_{ij} = S_0 \times \left(1 + \frac{S' \times (2i - m - 1)}{m - 1}\right) \text{ if } i = j. \tag{2}$$

In the model, we refer to B' and S' as the bias gradient and sensitivity gradient respectively. When fitted to data, B' and/or S' were free parameters for individual observers, and if not fitted then B'=0 and/or S'=0. S_0 was also a free parameter for each observer, controlling the overall level of performance.

For the similarity simulations and modelling, we assumed that the similarity between Sloan letters arises solely from the structure of the letters and that the similarity between pairs of Sloan letters can be quantified by the Optotype Correlation (OC) matrix shown in Fig. 5 (Fülep et al., 2017). The calculation is based on Pearson's normalized cross-correlation of the original letters (Neto et al., 2013). To simplify the model, we proposed that similarity is symmetrical (e.g., if H is similar to K, then K is similar to H by the same amount). We also propose that the pattern of the similarity between letters does not change for different letter sizes.

Another key assumption is that, while keeping the similarity pattern unchanged, the overall strength of the similarity between letters can vary from observer to observer. This is controlled by multiplying the off-diagonal correlation values by a factor that we call the Confusion

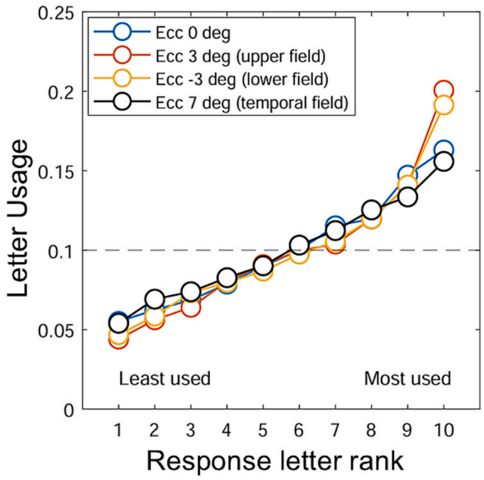

Fig. 4. Letter usage in the current experiment (Ecc 7 deg) and our previous one (Ecc 0, 3 and −3 deg) (Georgeson et al., 2023). Letter usage is expressed as the proportion of responses, on which each of the 10 letters was reported (correctly or not), averaged over the letter sizes and observers. The proportions of responses were rank-ordered separately for each observer, then averaged over observers. Note that the gradient of letter usage in the current experiment is consistent with that of the previous work.

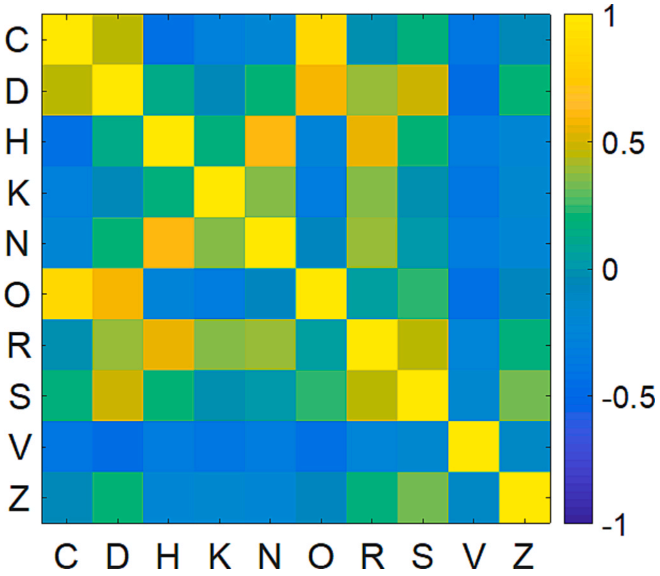

Fig. 5. The optotype correlation matrix (OC) for Sloan letters. A correlation value of one indicates identical letters, zero means no correlation, and a negative correlation indicates dissimilar letters. A higher positive value means stronger similarity between the corresponding letters.

Strength (Cs). Using the OC values and the Cs factor to quantify the similarity between Sloan letters for a given observer, the sensitivity of the i^{th} template to the j^{th} test letter ($i \neq j$) is:

$$S_{ij} = \left(S_0 \times Cs \times OC_{ij}\right) \text{ if } i \neq j \tag{3}$$

where Cs is a free parameter. Thus, the sensitivity of the i^{th} template when $i \neq j$ is controlled by its correlation with the j^{th} letter (i.e., OC_{ij}, see Fig. 5) and by Cs for a given observer. For example, consider the test letter C and suppose $S_0 = 1$ and $Cs = 1$. In this case, the sensitivity of the letter C template to the test letter C is 1, while sensitivity of the letter O template is 0.866, but that of the (rather dissimilar) letter H template is -0.408. Similar considerations apply to the rest of the templates.

Letter identification improves with increasing letter size. We capture this essential behaviour by assuming that the template's mean response (μ_{ij}) increases as a function of sensitivity S_{ij}, bias B_i, and test letter size t (expressed as letter stroke width, in min arc)

$$\mu_{ij} = B_i + sign(S_{ij} \times t).abs(S_{ij} \times t)^p \tag{4}$$

where p is an exponent of a power function relation between the mean response μ and letter size t. The exponent p controls the slope of the underlying psychometric function (proportion of correct responses *vs* test letter size). In our previous model p was a constant, but here the exponent p was a free parameter for each observer. We adopted the standard SDT assumption that the i^{th} template output is perturbed by additive zero-mean Gaussian noise with variance σ_i^2, and that $\sigma_i = 1$ for all i. The decision rule for letter identification was the MAX operator, i.e., the template that has the highest activity is chosen as the response letter (Fig. 2A).

3.3. Letter identification

In the modelling of bias and sensitivity, a more positive μ_{ij} increases the likelihood that the corresponding letter will generate the MAX response (Fig. 3B & C). On the other hand, in the similarity modelling, when a test letter is presented, the mean response becomes more positive for the test letter template but also for similar letters. They will be more likely to generate the MAX response across templates and tend to be *confused* with the test letter. For letters dissimilar to the test letter, the templates' mean responses shift in the opposite direction, making them less likely to be confused with the test letter. For example, similar letters like S and D are likely to be confused with each other, whereas dissimilar letters S and V are unlikely to be confused (Fig. 3D). Note that in Eq. (3), if Cs is zero, S_{ij} becomes zero when $i \neq j$. This means that our original NTM (Georgeson et al., 2023) is a special case of the present model, with zero Cs, hence orthogonal templates.

The *MATLAB* function used to compute (for all i, j) the probability P_{ij} that the i^{th} template delivers a response to the j^{th} test letter that is larger than that of all the other templates was taken from Zhou et al. (2014). For a given set of parameters (S_0, p, B', S', Cs), we used this function to compute the model's CM of frequency counts for each possible pairing of test letter *vs* response letter. In the experiment, we obtained one such 10 × 10 CM (test letters *vs* response letters) for each observer and each tested letter size.

3.4. Illustration of the effect of bias, sensitivity, and similarity on the letter identification task

To illustrate the expected effects of bias, sensitivity variation, and similarity between letters on the response pattern for individual letters, we ran simulations of letter identification tasks. We set $S_0 = 0.5$ and $p = 1.5$ to obtain a reasonable performance for the same five letter sizes tested in the current experiment. Four simulations were performed to collect model responses illustrating (i) bias ($B' = 1$, $S' = 0$ and $Cs = 0$), (ii) sensitivity variation ($B' = 0$, $S' = 1$ and $Cs = 0$), (iii) similarity between letters ($B' = 0$, $S' = 0$ and $Cs = 1$) or (iv) none of those factors ($B' = 0$, $S' = 0$ and $Cs = 0$). In case (i) bias was simulated with gradient $B' = 1$ (chosen for illustration). The calculated biases for individual letters (using Eq. (1) ranged from -1 to 1 and were assigned to the letter templates V, N, C, O, H, K, S, D, R and Z (arbitrarily ordered for demonstration), respectively. In this case, the bias of the letter V template is -1 (i.e., the largest bias against the letter) and that of the letter Z is 1 (i.e., the largest bias towards the letter). In case (ii), the sensitivity of each letter template when its preferred letter is presented (i.e., $i = j$) was simulated with gradient $S' = 1$. The calculated sensitivity for individual letters ranged from 0 to 1 (calculated using Eq. (2) and was assigned to the same order of letters V, N, C, O, H, K, S, D, R and Z, respectively. In this case, the V template has the lowest sensitivity while the Z template has the highest sensitivity towards its preferred letter. In case (iii), the influence of similarity between letters was simulated with $Cs = 1$. The sensitivity of each letter template to the non-preferred letter presented (i.e., $i \neq j$, hence similarity) is calculated using Eq. (3).

3.4.1. *Definitions of letter usage, correct usage and error usage*

In our previous work on letter biases (Georgeson et al, 2023) we defined a response measure (*letter usage*, LU_i) that is the unconditional probability of choosing the i^{th} response letter, whether correct or not. It can be estimated from the CM and Fig. 6 (left) illustrates this definition graphically, where the i^{th} response letter is 'R'. We now define two further measures, namely *correct usage* (*CU*) and *error usage* (*EU*) – again concerned with the profile of usage across letters, but now restricted to either correct or incorrect responses alone. *CU* and *EU* are conditional probabilities, and their definitions and relation to *LU* are described more fully in Appendix 2. Fig. 6 (center, right) illustrates their relation to the CM.

Having defined the three measures of response-letter usage (*LU, CU, EU*) we can now turn to the simulations. In the absence of bias, sensitivity variation and similarity between letters (Fig. 7, None), the model profiles of usage across letters were all flat (Fig. 7B,C,D, None). There was no over-calling or under-calling of any of the letters, either in general (*LU*), or in the distribution of correct (*CU*) or incorrect (*EU*) responses across letters.

However, in the bias scenario (Fig. 7, Bias), the biases added to the letter templates shifted the corresponding *LU* systematically (Fig. 7B, Bias) due to the shifted corresponding *CU* and *EU*. For example, the simulation shows that the letter Z (assigned the largest bias towards the letter) is over-called (either correctly or not) as shown by the corresponding high *CU* and *EU*. Consequently, the letter Z showed the highest *LU*. The opposite is observed for the letter with the largest negative bias (i.e., V) (Fig. 7C&D, Bias).

In the sensitivity simulation (Fig. 7, Sensitivity), the sensitivity differences between letters mainly affect the *CU* (Fig. 7C, Sensitivity). In this case, over-calling a given letter (e.g., Z) is due to the increased sensitivity towards that letter when presented (i.e., high *CU*) compared to the other letters. Letter V that is assigned with the lowest sensitivity to its preferred letter showed the lowest *CU* compared to other letters and was the least used.

The fourth column in Fig. 7 shows model performance with only similarity incorporated. Similarity between letters had minimal effect on overall letter usage (*LU*) (Fig. 7B, Similarity) compared with the effects of bias and sensitivity. This can be explained by the approximate averaging that occurs between *CU* and *EU*. For instance, the similarity between two frequently confused letters (e.g., C *vs* O) reduces the *CU* and increases the *EU* of both letters (Fig. 7C&D, Similarity), giving little change in *LU*. Additionally, the more dissimilar a letter is from the rest (e.g., V), the higher the *CU* and the lower *EU* for that letter (Fig. 7C&D, Similarity).

3.5. *Model fitting*

For each observer in the experiment, we obtained a 10 × 10 (presented *vs* responded) CM for each of the five letter sizes. The model was fitted to the observed number of responses (in each entry of the 500 response counts [i.e. 100 cells for each of the 5 confusion matrices]), using maximum likelihood – adjusting parameter values S_0, p, B', S', and Cs to maximize the log-likelihood (*LL*) of the parameters given the data. To calculate the *LL*, we used the following equation:

$$LL = \sum\{n_{ijk} \times log(P_{ijk})\} \tag{5}$$

where n_{ijk} is the number of choices of the i^{th} letter made when the j^{th} letter was presented at the k^{th} letter size. P_{ijk} is the model's proportion of responses for the corresponding ij pair at the k^{th} size. The variables n and P range over 100 letter pair combinations for each letter size, and the summation takes place over those 500 pairs.

We fitted eight versions of the template model, in which the fitted parameters represented B', S' & Cs, denoted as the model (B1C1S1), or B' & Cs (B1C1S0), B' & S' (B1C0 S1), Cs & S' (B0C1S1), B' only (B1C0S0), Cs only (B0C1S0), S' only (B0C0S1) or none (B0C0S0). The fitting of the B1C1S1 model was performed in two steps. Firstly, we adjusted overall sensitivity S_0 and slope (p) in a 2D sampled grid search to find the best-fitting value that maximised *LL* assuming that there is no bias ($B' = 0$), no sensitivity differences ($S' = 0$), and no similarity between letters ($Cs = 0$). Secondly, while using the best S_0 and p we ran a 3D sampled grid search to find the best values of the B', Cs, and S'. Then we readjusted for S_0 and p using the best B', Cs, and S'. We repeated the procedure until there was no change in the estimated parameters.

A similar approach was used to fit the B1C1S0, B1C0S1, and B0C1S1 models. In these cases, a 2D sampled grid search was performed first to fit for S_0 and p, followed by 2D sampled grid search to fit for either B' & Cs, B' & S' or Cs & S'. We repeated the procedure until there was no change in the estimated parameters. For the remaining models, a 3D sampled grid search was performed to fit the B1C0S0, B0C1S0, and B0C0S1 models to find the best values of the (S_0, p & B'), (S_0, p & Cs) or (S_0, p & S') respectively. A 2D sampled grid search was performed to fit the B0C0S0 model to find the best value of the S_0 and p. All the fittings were performed for each observer separately.

In the following section, we present the results of fitting the eight

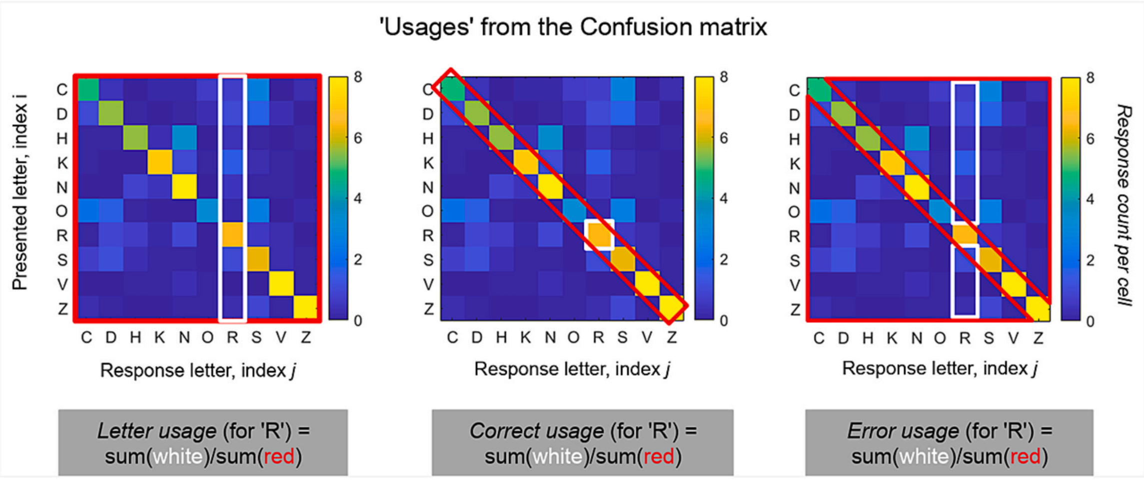

Fig. 6. Illustrates the empirical definitions of letter usage (*LU*, left), Correct Usage (*CU*, center) and Error Usage (*EU*, right). These measures are related to unconditional and conditional probabilities of letter choices described in Appendix 2.

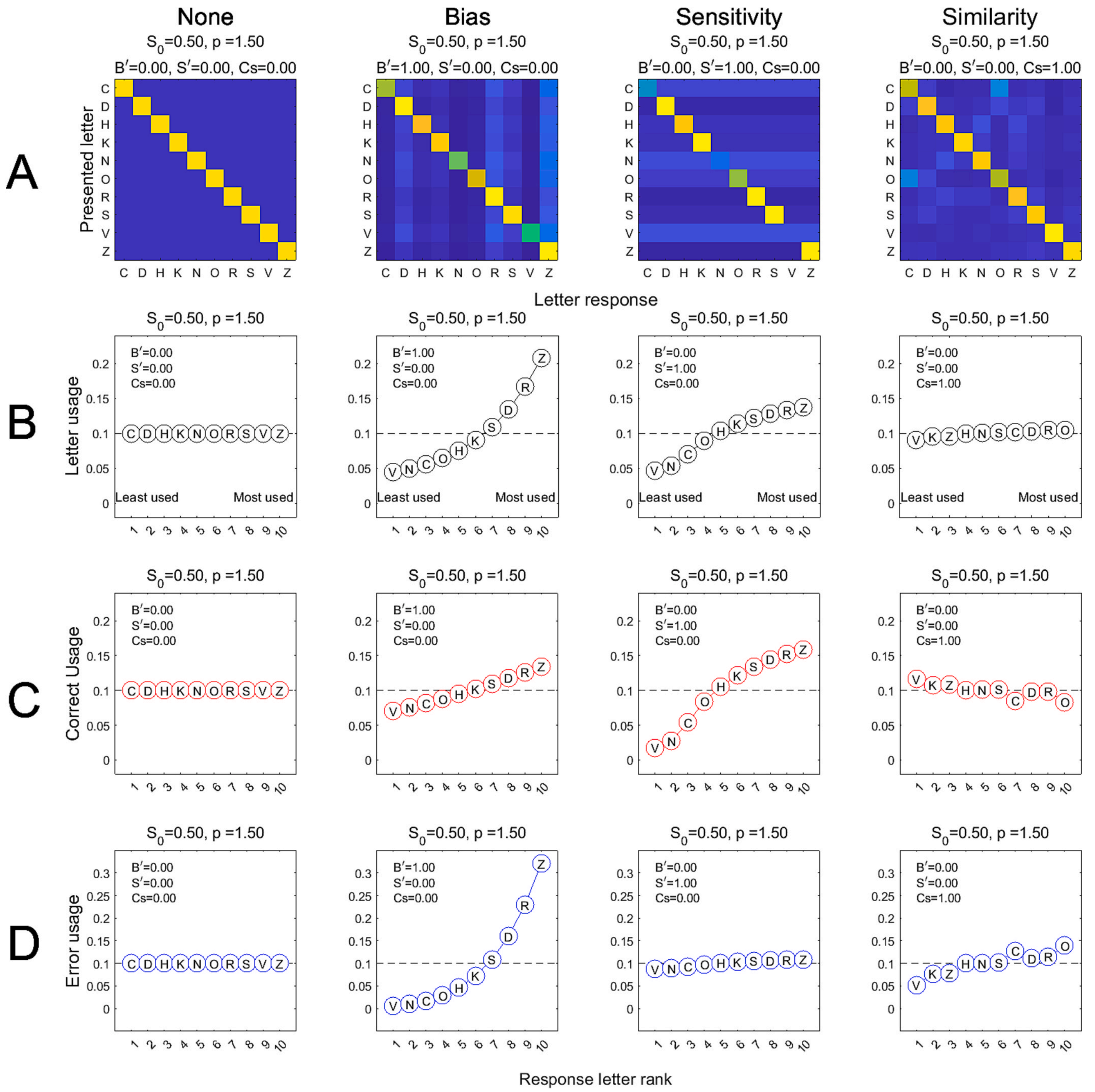

Fig. 7. Simulation of the effect of bias, sensitivity, similarity or none on the usage of letters in the identification task. A) the CMs of the simulated data pooled across letter sizes, B) Letter Usage, C) Correct Usage and D) Error Usage for each case. The dashed horizontal line is the usage (0.1) expected when all response letters are used equally often.

models to the experimental data and the comparison between them using the Akaike Information Criterion (AIC, Akaike, 1974) to determine the best-fitting model.

4. Results

4.1. Initial observations

The experimental data are summarized in Figs. 8 and 9. Fig. 8A shows the CM (presented letter *vs* letter response) collapsed across observers and letter sizes. The *LU*, *CU* and *EU* in Fig. 8B,C&D respectively, show a characteristic trend in proportions of usage across letters. Through simulation, we demonstrated (Fig. 7) that a similar systematic shift in usage occurred when incorporating bias and sensitivity in the task (see “Illustration of the effect of bias, sensitivity and similarity on letter identification task” in methods).

Fig. 9A shows the CM separately for each letter size averaged across observers. Note that the pattern of responses at larger letter sizes (e.g., letter size #3) resembles the pattern of the OC matrix (Fig. 5), suggesting that the performance is also influenced by the similarity between letters. Since the data here were averaged across observers, and each observer might show a different preference in biases or sensitivities towards letters, the effect would be diminished by averaging. Fig. 9B shows the CM recalculated such that both the stimulus letter set (vertical axis) and the

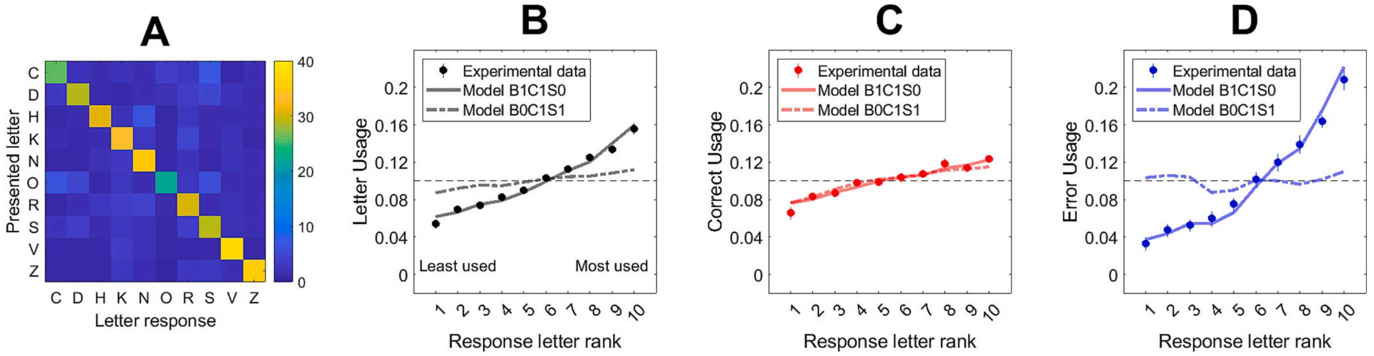

Fig. 8. Shows the experimental data as A) CM (presented letter *vs* letter response). Data were collapsed across letter sizes first and then averaged across observers. Here and throughout (unless mentioned otherwise), the colour scale illustrates the frequency count of letter responses, where warmer colours show higher frequencies. The diagonal cells represent correct responses, whereas the non-diagonal cells represent incorrect responses. The figure also shows the distribution of B) Letter Usage, C) Correct Usage and D) Error Usage across response letters of the experimental data (±SE) and two selected models (B1C1S0 and B0C1S1). Note how model B1C1S0 (i.e., bias and similarity), in contrast to model B0C1S1 (i.e., sensitivity and similarity), closely captures the systematic shift observed across the three types of usage. The letters were rank-ordered according to their Letter Usage for each observer from the least to the most used letters then averaged. The Correct and Error usage follow the same letter order as the Letter Usage.

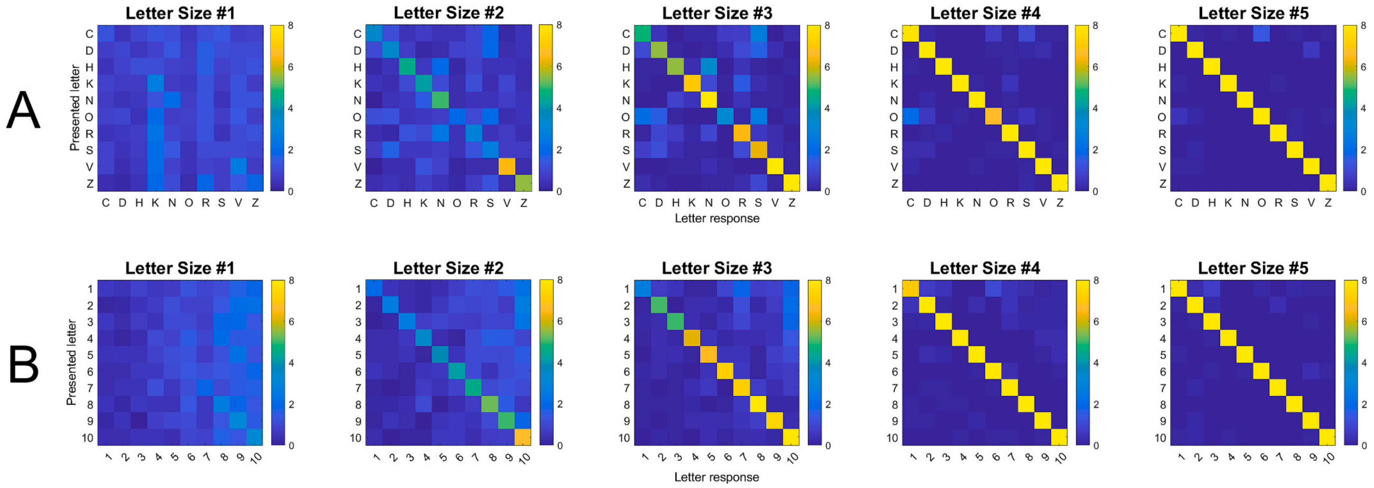

Fig. 9. Shows a) the experimental data presented as the cm separately for each letter size, but averaged across observers. b) shows the cm recalculated such that both the stimulus letter set (vertical axis) and the response letter set (horizontal axis) were reordered, for each observer, according to their rank order of lu from least to most frequently used letters. the cms were then averaged across observers. Note that CMs in B are for individual letter sizes, but the ranking by Letter Usage was done after pooling across all sizes, separately for each observer.

response letter set (horizontal axis) were reordered, for each observer, according to their rank order of LU from least to most frequently used letters. The CM were then averaged across observers. This reveals that the observers clearly report some letters more than others, especially at the two smallest letter sizes. Note that at the two smallest letter sizes, where there is high uncertainty in the performance, the responses are determined mainly by the bias which (in our model) is, unlike similarity and sensitivity, independent of the sensory input strength (i.e., letter size). As the letter size increases, the effect of similarity (frequency of responses of certain letter pairs) and/or sensitivity (frequency of the correct responses, i.e., diagonal cells) become more prominent (Fig. 9A). Different model variants were fitted to the experimental data to reveal the effect of the three factors (i.e., bias, sensitivity and/or similarity) in the letter identification task in the current experiment (Fig. 10).

Fig. 10A shows the resulting confusion matrices of the fitted models averaged across the five letter sizes and then averaged across observers, where each column shows the fit of one model (the basic model B0C0S0 is not included). To facilitate the identification of the model that accounts best for the experimental CM data (Fig. 8A), the difference between each model and the experimental data was calculated for each subject using the χ^2 value and these values were then summed across subjects. For each model (Fig. 10A), χ^2 value was calculated as:

$$\chi^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{6}$$

where O_{ij} is the frequency count of observed responses for cell ij in the experimental CM and E_{ij} is the expected frequency of responses for the corresponding cell ij in the model CM. A constant of 1 was added to all matrices to avoid division by zero.

Models B1C1S1 and B1C1S0 show the lowest χ^2 values of 1108.17 and 1179.47, respectively. Models B1C1S1 and B1C1S0 also give a good account of the *CU* and *EU* at individual and group levels (Fig. 10B&C). Fig. 10 suggests that bias is the most influential factor in model B1C1S1 and B1C1S0 since it gives the smallest χ^2 value (1503.03) when fitted individually (B1C0S0), compared with the sensitivity-only model (B0C0S1) where χ^2 is higher (1977.55), and when compared with similarity-only (B0C1S0) where χ^2 is again higher (2203.45) (Fig. 10A).

Furthermore, bias is the key factor that accounts for some of the *CU* and most (if not all) of the *EU* when fitted individually (B1C0S0) or in conjunction with the other factors (B1C1S0, B1C0S1, and B1C1S1) (Fig. 10B&C). Since the χ^2 values were calculated for the average across

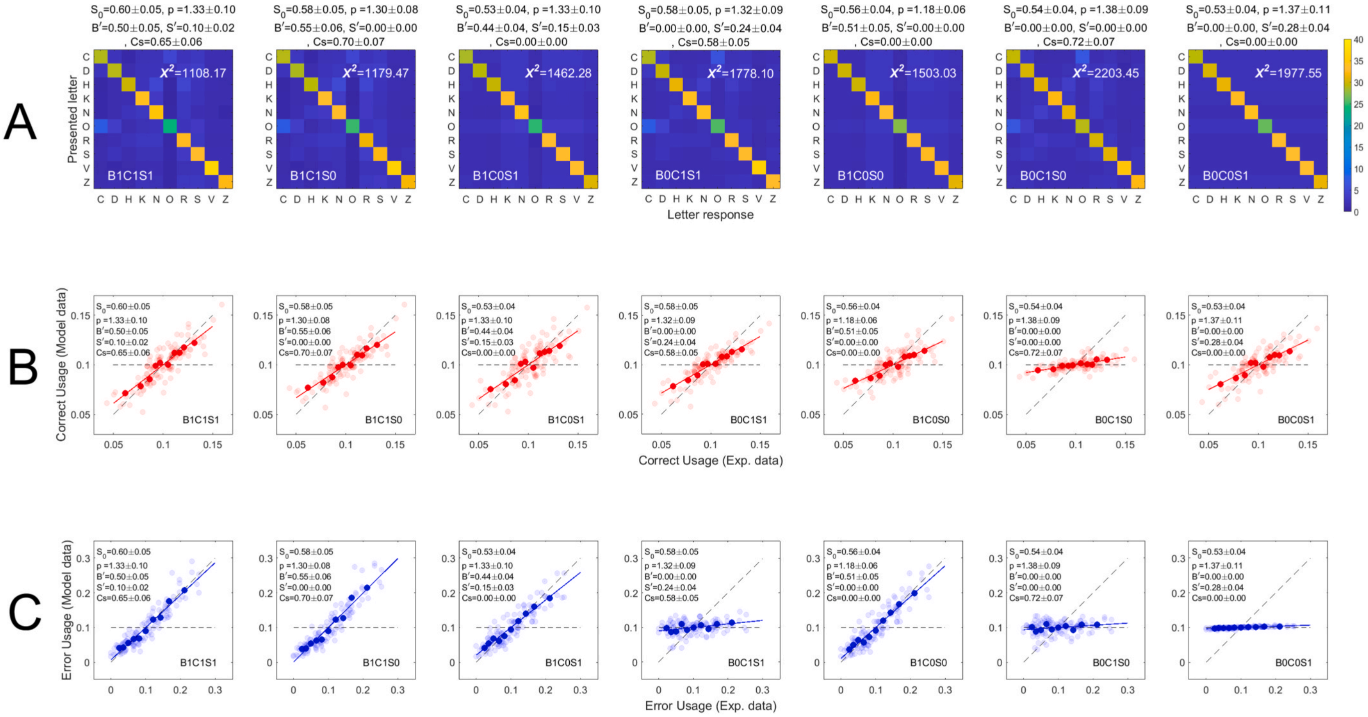

Fig. 10. Shows seven variants of the model after excluding the basic model b0c0s0. a) for each model, the cm is averaged across the five letter sizes for each observer and then averaged across observers.χ^2 value in each matrix quantifies the difference between a given model and the experimental data (after rank-ordering the data, see text). The smaller the value, the better the model is in accounting for the experimental CM data. B) shows the agreement between the model and experimental data in Correct Usage. The faint red points are the observers' data for individual letters. For each observer, the letters' Correct Usage values from experimental data were rank-ordered first (from low to high). Then these ranked *CU* values were averaged across observers. The model average *CU* was calculated across each observer's model *CU*s using the same rank ordering as the experimental *CU*s. The solid red data points (and the fitted red line) show the agreement between the two averages. C) The same procedure was applied to calculate and depict (in blue) the averages for Error Usage. Here and throughout (unless mentioned otherwise), the diagonal dashed line represents perfect agreement between model and experimental usages, while data points on the horizontal dashed line would represent the same model usage (CU or EU) across all response letters, quite unlike the experiment.

the rank-ordered data, the results here might underestimate the contribution of the similarity factor. However, Fig. 10B&C show that the similarity factor alone (B0C1S0) accounted poorly for either *CU* or *EU* when compared with sensitivity-alone (B0C1S0) or bias-alone (B1C0S0).

4.2. Formal model comparison

To assess these observations about the models, we conducted a formal model comparison using the *AIC* (Akaike Information Criterion; Akaike, 1974) at the group level (Table 1). Akaike weights are derived for each model in the set, and represent the probability that a given model is the best, most plausible model of those considered in the set (Burnham & Anderson, 2002, p.75). The weight was 0.79 for the fuller model B1C1S1, followed by model B1C1S0 with a weight of 0.21. Therefore, the fuller model B1C1S1 is the best or most-favoured model among the eight candidates (Table 1).

This finding was confirmed by the analysis of deviance for pairs of nested models (Collett, 2003, p. 73). The analysis of deviance tests whether addition of an extra parameter to a model produces a significant improvement in fit that justifies the extra flexibility offered by that parameter. Here we found that bias (B0C1S1 *vs* B1C1S1, $\chi^2 = 816.42$, $df = 12$, $p < 0.001$), sensitivity (B1C1S0 *vs* B1C1S1, $\chi^2 = 27.04$, $df = 12$, $p = 0.008$), and similarity factors (B1C0S1 *vs* B1C1S1, $\chi^2 = 674.64$, $df = 12$, $p < 0.001$) all had statistically significant roles to play in the fitting of the fuller model B1C1S1.

Further AIC model comparisons made after excluding the fuller model (B1C1S1) revealed that the bias and similarity model (B1C1S0) was favoured over the other six models (Table 2). Additionally, comparing only the one-factor models revealed that the bias model (B1C0S0) was favoured over the other two models (B0C1S0 & B0C0S1) (Table 3). Comparing the similarity model to the sensitivity model after excluding the bias model showed that the similarity model was favoured over the sensitivity model (Table 4).

These findings suggest that (i) bias has a higher contribution to the fuller model than the other two factors (sensitivity and similarity) and (ii) the joint effect of bias and similarity is more influential in the fuller

Table 1
Comparison of the eight models via AIC analysis, for the group of 12 observers.

Model	*LL*	*K*	*n*	*AIC*	*Δ AIC*	*Ak. Wt*
B0C0S0	−21493.34	24	6000	43034.87	1720.93	0
B0C0S1	−21282.15	36	6000	42636.75	1322.81	0
B0C1S0	−21158.59	36	6000	42389.62	1075.68	0
B1C0S0	−20980.90	36	6000	42034.25	720.30	0
B0C1S1	−21004.57	48	6000	42105.94	791.99	0
B1C0S1	−20933.68	48	6000	41964.16	650.21	0
B1C1S0	−20609.88	48	6000	41316.54	2.60	0.21
B1C1S1	−20596.36	60	6000	41313.94	0	0.79

Table 2
Comparison of seven models via AIC analysis, for the group of 12 observers.

Model	*LL*	*K*	*n*	*AIC*	*Δ AIC*	*Ak.Wt*
B0C0S0	−21493.34	24	6000	43034.87	1718.33	0
B0C0S1	−21282.15	36	6000	42636.75	1320.21	0
B0C1S0	−21158.59	36	6000	42389.62	1073.08	0
B1C0S0	−20980.90	36	6000	42034.25	717.70	0
B0C1S1	−21004.57	48	6000	42105.94	789.40	0
B1C0S1	−20933.68	48	6000	41964.16	647.61	0
B1C1S0	−20609.88	48	6000	41316.54	0	1

Table 3
Comparison of four models via AIC analysis, for the group of 12 observers.

Model	*LL*	*K*	*n*	*AIC*	*Δ AIC*	*Ak.Wt*
B0C0S0	−21493.34	24	6000	43034.87	1000.62	0
B0C0S1	−21282.15	36	6000	42636.75	602.50	0
B0C1S0	−21158.59	36	6000	42389.62	355.38	0
B1C0S0	−20980.90	36	6000	42034.25	0	1

Here and throughout: *LL*, total log likelihood; *K*, total no. of free parameters across the group; *n*, total no. of data points; *AIC* is the Akaike score; *Δ AIC* is the difference between each model's *AIC* value and the lowest value in the set of scores; *Ak.Wt* refers to the Akaike weights.

Table 4
Comparison of three models via AIC analysis, for the group of 12 observers.

Model	*LL*	*K*	*n*	*AIC*	*Δ AIC*	*Ak.Wt*
B0C0S0	−21493.34	24	6000	43034.87	645.25	0
B0C0S1	−21282.15	36	6000	42636.75	247.13	0
B0C1S0	−21158.59	36	6000	42389.62	0	1

model than the joint effect of either bias and sensitivity or similarity and sensitivity. The results also suggest (iii) that the sensitivity factor has the least effect on the fuller model.

Our results show that the best model (B1C1S1) accounts very well for the variation in usage across letters for incorrect responses (*EU*) (Fig. 10C). However, we also investigated more specifically how the best model accounts for the *sources* of differences in incorrect response usages which can be either bias, similarity between letters, or both. We employed Luce's choice model (Luce, 1963) to compute bias and similarity from the best model B1C1S1 and the experimental data, and then examined the agreement between the outcomes of the model and experimental data (Fig. 11). A custom-made *MATLAB* function to compute the response biases and similarity parameters using Luce's choice model was employed and is freely accessible here: https://github.com/HBarhoom/Codes-.

The fitting of Luce's choice model showed that model B1C1S1 described the bias and similarity in the experimental data very well. Fig. 11A shows that Luce's bias in the best model is strikingly similar to Luce's bias in the experimental data (coefficient of determination, $R^2 = 0.94$). This also indicates that our model estimates the same kind of bias that Luce's model does, and it supports our strong assumption about the linearity of the bias and its ranking based on *LU* (Georgeson et al, 2023). Fig. 11B shows Luce's similarity parameters for the 45 confusion pairs. There are only 45 confusion pairs because Luce's and our model assume that the similarity between letters is symmetrical (i.e., the parameter of similarity between the presented letter C and the responded letter O is identical to the similarity parameter between the presented letter O and the responded letter C). The results suggest that Luce's similarity between letters estimated from the best model data is similar to Luce's similarity from experimental data. It is also evident that the agreement (Fig. 11B) is not as strong as in the case of Luce's bias comparison (Fig. 11A) ($R^2 = 0.73$). This is not unexpected since we used the *OC* matrix to capture the similarity between letters that follows the pattern of the *OC* matrix. Hence, Luce's model will capture a particular pattern of similarity when fitted to the best model data (i.e., the pattern of *OC* matrix) whereas Luce's model captures a variety of similarity patterns for each observer when fitted to experimental data. Nevertheless, the reasonable agreement between Luce's similarity of the best model and experimental data suggests that the general and common similarity pattern in experimental data is the one that follows the *OC* matrix (Fig. 5).

5. Discussion

Georgeson et al. (2023) recently introduced a model to compare the influence of biases and sensitivities of individual letters in letter identification. Here we extended this model by incorporating the similarity between letters in the NTM to investigate the joint role of the three factors – bias, sensitivity and similarity – rather than bias *vs* sensitivity only.

The results clearly show that the model with all three factors (i.e., B1C1S1) was the *best* model and gave an excellent account of the distribution of total, correct and incorrect responses in the experimental data (Fig. 10, Table 1). Results also showed that biases were the major source of errors in the task and contributed to a greater degree than either similarity or sensitivity to the variation in correct responses across letters (Fig. 10B and C).

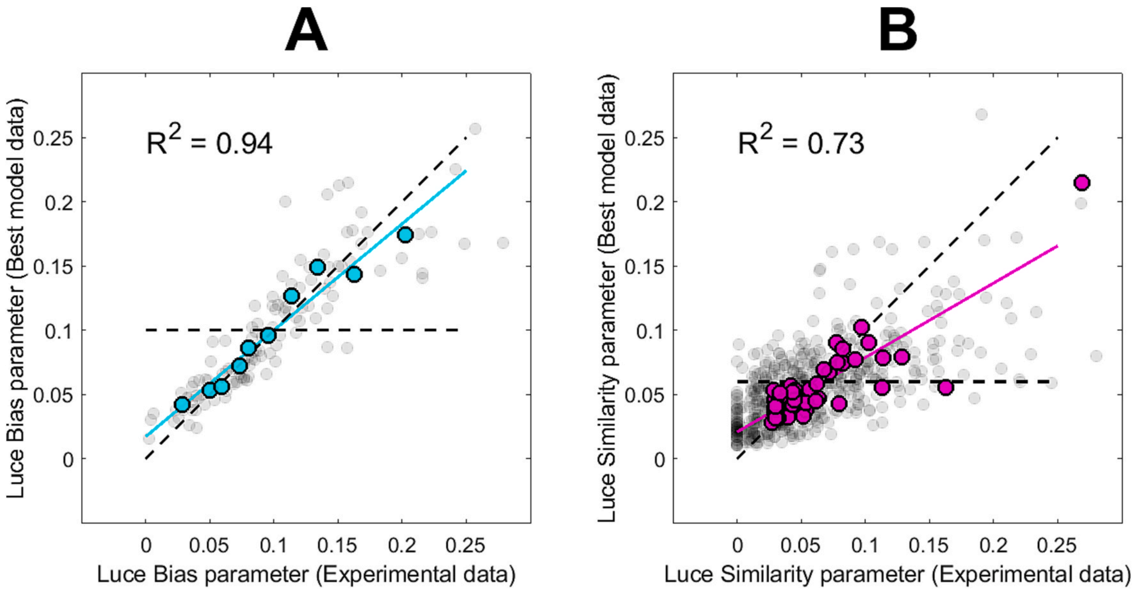

Fig. 11. Shows a) the agreement between luce's bias parameters estimated from the best model and experimental data. the faint grey data points are the individual bias parameters of each letter (10 letters) for each subject and the cyan data points show the averages across subjects, with the cyan line indicating the fitted regression. b) shows the agreement between luce's similarity parameters of the model and experimental data. the faint grey data points are the individual similarity parameters of each pair (45 pairs) for each subject and the magenta data points are the average across subjects, with the magenta line indicating the fitted regression.

In our model for bias, we assume that the biases for individual letters can be rank-ordered based on the overall *LU* of these individual letters when pooled across the letter sizes. The rank-order is unique for each observer. This assumption has been investigated extensively in our previous work and the experimental data supports such a systematic order (Georgeson et al., 2023).

For the modelling of similarity, we assumed that the patterns of similarity between letters for a given observer must follow the *OC* matrix pattern. It seemed unlikely that the pattern of similarity observed and captured by the model in the experimental data could arise as an artefact of random error (i.e., noise). But to make sure, we tested this in a simulation (see Appendix 1) and confirmed that random error alone was very unlikely to create an *apparent* similarity pattern ($p < 0.0001$).

Luce’s choice model (Luce, 1963) has been shown to perform very well in capturing response bias and similarity in letter identification tasks (Nosofsky, 1991; Smith, 1992; Mueller & Weidemann, 2012; Coates, 2015; Hamm et al., 2018). Here we employed Luce’s choice model to further validate our model. We investigated whether the best-performing model (B1C1S1) is efficient in capturing bias and similarity by calculating and comparing bias and similarity parameters from model B1C1S1 and the experimental data using Luce’s choice model. Results show that model B1C1S1 was remarkably effective (especially for the bias) in capturing the bias and similarity as shown by the agreement in Luce’s bias and similarity parameters estimated from model B1C1S1 and experimental data (Fig. 11).

In the current study, we examined the fitting of the proportions of responses of the model to the corresponding proportions of responses in the experimental data (either correct or incorrect). This means that we imposed a high restriction on the model fitting such that a good fit to the correct and incorrect response proportions in the experimental data would support our strong assumptions regarding letter usage-based rank-order and *OC* pattern of similarity. Interestingly, the model with these constraints gave an excellent account of the experimental data, thus giving strong support to those model assumptions.

The current study might underestimate the effect of similarity and its contribution to the *best* model's fit. This is because modelling the bias and similarity are essentially different: modelling bias used some information from the experimental data (i.e., *LU* rank-order) while the *pattern* of similarity modelling was entirely objective. We used the same *OC* matrix as an objective source of information about inter-letter similarity, modulated by the Cs factor derived for individual subjects.

In practice, one might also expect to find individual differences in *OC* patterns that could improve the estimation of similarity. Hence in future we may be able to improve the structure and fit of the similarity model by including information about individual differences in *OC* patterns. Nevertheless, we conclude that incorporating the similarity factor into the NTM (with the current assumptions) improved our understanding of the simultaneous contribution of the bias, sensitivity and similarity between letters in the letter identification task.

CRediT authorship contribution statement

Hatem Barhoom: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mark A. Georgeson:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization. **Mahesh R. Joshi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Conceptualization. **Gunnar Schmidtmann:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Conceptualization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2026.108822.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Xplore: IEEE Transactions on Automatic Control, 19*(6), 716–723.

Alexander, K. R., Xie, W., & Derlacki, D. J. (1997). Visual acuity and contrast sensitivity for individual Sloan letters. *Vision research, 37*(6), 813–819.

Anderson, R., & Thibos, L. (2004). The filtered Fourier difference spectrum predicts psychophysical letter discrimination in the peripheral retina. *Spatial vision, 17*(1–2), 5–15.

Barhoom, H., Joshi, M. R., & Schmidtmann, G. (2021). The effect of response biases on resolution thresholds of Sloan letters in central and paracentral vision. *Vision Research, 187*, 110–119.

Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision research, 11* (5), 459–474.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision, 10*(4), 433–436.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed). New York, NY: Springer, New York.

Coates, D. R. (2015). Quantifying crowded and uncrowded letter recognition (Doctoral dissertation, University of California, Berkeley).

Collett, D. (2003). *Modelling Binary Data* (((2nd ed.).).). Boca Raton & London: Chapman & Hall/CRC.

Ferris, F. L., III, Kassoff, A., Bresnick, G. H., & Bailey, I. (1982). New visual acuity charts for clinical research. *American journal of ophthalmology, 94*(1), 91–96.

Fiset, D., Blais, C., Ethier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for identification of uppercase and lowercase letters. *Psychological science, 19*(11), 1161–1168.

Fülep, C., Kovács, I., Kránitz, K., & Erdei, G. (2017). Correlation-based evaluation of visual performance to reduce the statistical error of visual acuity. *JOSA A, 34*(7), 1255–1264.

Georgeson, M. A., Barhoom, H., Joshi, M. R., Artes, P. H., & Schmidtmann, G. (2023). Revealing the influence of bias in a letter acuity identification task: A noisy template model. *Vision Research, 208*, Article 108233.

Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics, 25*(5), Article 425431.

Graham, N. V. S. (1989). *Visual pattern analyzers*. Oxford University Press.

Grimm, W., Rassow, B., Wesemann, W., Saur, K., & Hilz, R. (1994). Correlation of optotypes with the Landolt ring—a fresh look at the comparability of optotypes. *Optometry and Vision Science, 71*, 6–13.

Hairol, M. I., Abd-Latif, N. A., Low, P., Lim, W. P., Aik, J. Y., & Kaur, S. (2015). Effects of foveal and eccentric viewing on the resolution and contrast thresholds of individual letters. *Psychology & Neuroscience, 8*(2), 183–192.

Hamm, L. M., Yeoman, J. P., Anstice, N., & Dakin, S. C. (2018). The Auckland Optotypes: An open-access pictogram set for measuring recognition acuity. *Jouranl of Vision, 18* (3), 13.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?.

The MathWorks. (2024). *MATLAB version 24.1.0 (R2024a)*.. Natick, Massachusetts: The MathWorks, Inc. https://www.mathworks.com.

“Detection and recognition”. Handbook of mathematical psychology (Vol. 1,, (1963), 103–189.

Ludvigh, E. (1941). Extrafoveal Visual Acuity as measured with Snellen Test-Letters*. *American Journal of Ophthalmology, 24*(3), 303–310.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and“ nonparametric” indexes. *Psychological bulletin, 107*(3), 401.

Macmillan, N. A., & Creelman, C. D. (2005). Detection Theory: a user’s guide (2nd ed.). Mahwah, NJ & London: Lawrence Erlbaum Associates.

Mueller, S. T., & Weidemann, C. T. (2012). Alphabetic letter identification: Effects of perceivability, similarity, and bias. *Acta Psychologica, 139*(1), 19–37.

Navarro, R. (2009). The optical design of the human eye: A critical review. *Journal of Optometry, 2*(1), 3–18.

Neto, A. M., Victorino, A. C., Fantoni, I., Zampieri, D. E., Ferreira, J. V., & Lima, D. A. (2013). Image processing using Pearson's correlation coefficient: Applications on autonomous robotics. 13th International Conference on Autonomous Robot Systems (16). IEEE.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology, 23*, 94–140.

Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*, 437–442.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision research, 46*(28), 4646–4674.

Reich, L. N., & Bedell, H. E. (2000). Relative legibility and confusions of letter acuity targets in the peripheral and central retina. *Optometry and Vision Science, 77*(5), 270-.

Shah, N., Dakin, S. C., & Anderson, R. S. (2012). Effect of optical defocus on detection and recognition of vanishing optotype letters in the fovea and periphery. *Investigative ophthalmology & visual science, 53*(11), 7063–7070.

Shah, N., Dakin, S. C., Redmond, T., & Anderson, R. S. (2011a). Vanishing Optotype Letter Acuity: Repeatability and effect of the Number of Alternatives. *Ophthal Physiol Opt, 31*, 17–22.

Shah, N., Laidlaw, D. A. H., Shah, S. P., Sivasubramaniam, S., Bunce, C., & Cousens, S. (2011b). Computerized repeating and averaging improve the test-retest variability of ETDRS visual acuity measurements: Implications for sensitivity and specificity. *Investigative ophthalmology & visual science, 52*(13), 9397–9402.

Sloan, L. L. (1959). New test Charts for the Measurement of Visual Acuity at far and Near Distances*. *American Journal of Ophthalmology, 48*(6), 807–813.

Smith, J. E. K. (1992). Alternative biased choice models. *Mathematical Social Sciences, 23*, 199–219.

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision, 11*(5), 13.

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. Perception & Psychophysics, 9, 40-50.

Townsend, J. T. (1971). Alphabetic confusion: A test of models for individuals. Perception & Psychophysics, 9, 449-454.

Zhou, J., Georgeson, M. A., & Hess, R. F. (2014). Linear binocular combination of responses to contrast modulation: Contrast-weighted summation in first-and second-order vision. Journal of vision, 14(13), 24-24.