Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-021-01943-8

BRIEF REPORT

Temporal crowding is a unique phenomenon reflecting impaired target encoding over large temporal intervals

Shira Tkacz-Domb[1,2] · Yaffa Yeshurun[1]

Accepted: 30 April 2021

Abstract

Crowding refers to impaired object identification when presented with other objects, and it is well established that spatial crowding—crowding from adjacent objects—affects many aspects of visual perception and cognition. A similar interference also occurs across time—the identification of a target object is impaired when distracting objects precede and succeed it. When such interference is observed with relatively long interitem intervals it is termed temporal crowding. Thus far, little was known about temporal crowding and its underlying processes. Particularly it was unknown which aspects of visual processing are impaired by temporal crowding, and the answer to this question bears critical theoretical implications. To reveal the nature of this impairment we used a continuous-report task and a mixture-model analysis. In three experiments, observers viewed sequences of three oriented items separated by relatively long intervals (170–475ms). The target was the second item in the sequence, and the task was to reproduce its orientation. The findings suggest that temporal crowding impairs target encoding and increases substitution errors, but there was no evidence of a reduced signal-to-noise ratio. This pattern of results was similar regardless of stimuli duration and target–distractor similarity. However, it differed considerably from the pattern found for ordinary masking and spatial crowding, indicating that temporal crowding is a unique phenomenon. Moreover, the finding that temporal crowding affected the precision of target encoding even when the items were separated by almost half a second suggests that visual processing requires a surprisingly long time to complete.

Keywords Visual crowding · Temporal processing · Masking · Statistical mixture models · Visual representation latency

The ability to identify a target is lower when other items are also present than when it is presented in isolation; a phenomenon termed 'crowding'. Crowding was studied extensively in the spatial domain—when the target is flanked by other stimuli that are presented simultaneously with the target (e.g., Bouma, 1970; Ester et al., 2014; Rashal & Yeshurun, 2014; Strasburger, 2005; Yeshurun & Rashal, 2010; for a review see Whitney & Levi, 2011). Crowding occurs also in the time domain—when target identification is impaired by other stimuli that surround it in time, but only a handful of studies (Bonneh et al., 2007;Tkacz-Domb & Yeshurun, 2017 ; Yeshurun et al., 2015) examined temporal crowding in its 'uncontaminated' manifestation—when there is no involvement of spatial crowding (i.e., 'pure' temporal crowding—the distractors appear at the *same* location as the target but on *different* time points). Two recent studies (Tkacz-Domb & Yeshurun, 2017; Yeshurun et al., 2015) used letter stimuli and a four-alternative forced-choice task requiring identification of the target-letter's orientation. In the crowded condition, the target and distractors were separated by varying stimulus-onset asynchrony (SOA), but the target could also appear by itself (uncrowded condition). Temporal crowding emerged in both studies: identification deteriorated as the letters were closer in time (i.e., as the SOA was shorter), and this SOA effect persisted beyond the limits of ordinary masking (i.e., beyond SOAs of 100–150 ms; Breitmeyer, 1984; Breitmeyer & Ogmen, 2000, 2006; Enns & Di Lollo, 2000; Gorea, 1987; Enns, 2004). Indeed, performance in the crowded condition was worse than in the uncrowded condition even with SOAs

✉ Yaffa Yeshurun
yeshurun@research.haifa.ac.il

Shira Tkacz-Domb
shirtzi@yahoo.com

1 School of Psychological Sciences and Institute of Information Processing and Decision Making, University of Haifa, Haifa, Israel

2 Present address: Department of Electrical Engineering and Computer Science, York University, Toronto, ON M3J 1P3, Canada

Published online: 02 June 2021

longer than 400 ms. Moreover, temporal crowding was found even when there was no temporal uncertainty, and although it was reduced when attention was allocated to the target location it was not eliminated.

These long-lasting temporal effects could potentially challenge the common belief that the representation of visual information is established very fast, emerging around 150 ms, and that this is the case with both simple and complex scenes (e.g., Bankson et al., 2018; Greene & Oliva, 2009; Hung et al., 2005; Thorpe et al., 1996). However, because so little is known about temporal crowding, we need to better understand its characteristics and the processes underlying it, before we consider how current theories of visual perception may be modified to accommodate such long-lasting effects. For instance, if the impaired performance demonstrated in previous temporal crowding studies merely reflects source confusion—confusion between the representation of the target and that of the distractors, then this impairment has no bearing on the assumption that visual representation is generated very fast. However, if the impairment brought about by temporal crowding reflects, at least partially, a reduction in the quality of target representation then this would indeed suggest that generating a stable representation is considerably slower than the common belief.

To better understand the nature of the impairment brought about by temporal crowding, we used a continuous measure of perceived orientation. Specifically, the observer had to rotate a probe line to assume the perceived orientation of the target, and the measure of performance was response error—the difference between the target's orientation and that reproduced by the observer. The most commonly used model to analyze data obtained with such continuous-report tasks is a mixture model (e.g., Agaoglu et al., 2015; Asplund et al., 2014; Bays et al., 2009; Ester et al., 2014; Shechter & Yashar, 2021; Zhang & Luck, 2008). The model used here offers four parameters that describe the data: (1) The width (*sd*) of a Gaussian distribution of errors that is centered around the actual target orientation (i.e., zero error). This parameter reflects the error variance of trials in which the target was at least partially perceived. It conveys the precision of the encoding process or the quality of the representation; the smaller the *sd*, the higher the encoding precision. (2) The height (*g*) of a uniform distribution of errors that are the result of pure guessing. This parameter indicates the guessing rate. Because the guessing rate reflects the rate at which the target was not registered at all, Agaoglu et al. (2015) suggested that this parameter reflects the signal-to-noise ratio (SNR); the higher the *g*, the lower the SNR. (3–4) When a distractor is also present, another parameter is relevant—the rate of reporting the orientation of the distractor instead of the target (i.e., substitution errors). This is modeled by an additional Gaussian distribution centered on the distractor's orientation. Because our design included two distractors, we adopted the two-misreport mixture model (Shechter & Yashar, 2021), which allows a separate estimation of the rate of substitution errors for each of the two different distractors (β_1 & β_2).

In three experiments, we combined a continuous-report task with a temporal crowding paradigm. Specifically, a sequence of three randomly oriented stimuli was presented to the same location; the second stimulus in the sequence was the target (see Fig. 1). The target–distractors SOAs varied across trials, but all SOAs were longer than the temporal limits of ordinary masking. The sequence was followed by a probe, and the participants rotated it to reproduce the target's orientation. In an uncrowded condition, only the target was presented. We used the two-misreport mixture model to examine which of its parameters, and the processes they reflect, is affected by temporal crowding. In Experiment 1, all the stimuli were black and their duration was relatively long (75 ms). In Experiment 2, the stimuli were presented for a considerably shorter time (20 ms) to examine the role of stimuli duration. In Experiment 3, the target was black and the distractors were white to examine the role of target–distractors similarity. Lastly, we used the pattern of results that emerged from these experiments to compare temporal and spatial crowding as well as temporal crowding and ordinary masking.

General methods

Observers

Fifteen observers participated in Experiment 1, 17 observers participated in Experiment 2, and 15 observers participated in Experiment 3. Out of these overall numbers, two observers participated in all three experiments, one observer participated in Experiments 1 and 2, and four observers participated in Experiments 2 and 3. The sample size was based on previous successful demonstrations of temporal crowding (Yeshurun et al., 2015). Furthermore, power analysis obtained with the Shiny web app (Anderson et al., 2017) for a within-subject analysis of variance (ANOVA) using an uncertainty and publication bias correction, indicated that the minimum sample size required for the examination of SOA effects at a power >99 % with a Type I error ($\alpha < 0.05$) is nine participants. The F values, degrees of freedom, and effect sizes used in this analysis were based on Yeshurun et al. (2015; ISI effect), $F(9, 126) = 11.96$, $\eta_p^2 = 0.46$, $N = 16$. This analysis confirmed that the current study sample size had sufficient statistical power. All observers were students from the University of Haifa, with normal vision, who signed a consent form. Observers were naïve to the purpose of the study and were either paid or received course credit for participating. The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of the University of Haifa.

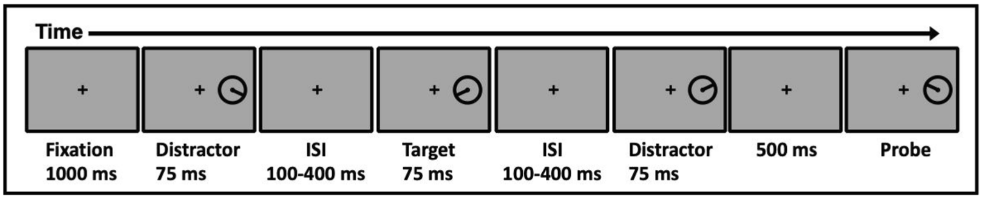

Fig. 1 A schematic depiction of the sequence of events in the crowded condition of Experiment 1. There were five possible target–distractors SOAs (175, 225, 275, 375, 475 ms). The SOA was constant within a trial, but varied between trials. In the uncrowded condition, the distractors were omitted. The task was to rotate the probe to reproduce the target orientation

Stimuli and apparatus

Stimuli were presented using the Psychophysics Toolbox extension (Brainard, 1997; Kleiner et al., 2007) in MATLAB on a 19-in. monitor of an IBM-compatible PC (1,024 × 768 resolution at a refresh rate of 85 Hz). Eye movements were recorded from the right eye with an EyeLink 1000 eye tracker (temporal resolution of 1000 Hz; SR Research, Ottawa, ON, Canada). A sequence of three stimuli was randomly presented to the right or the left of a central fixation circle (diameter 0.3°) at an eccentricity of 9° (see Fig. 1). The stimuli were separated by an SOA that varied between trials (Experiments 1 and 3: 175, 225, 275, 375, 475 ms; Experiment 2: 170, 220, 320, 420, 470 ms). The second stimulus in the sequence was defined as the target. Each stimulus consisted of a circle (diameter 2°) with an inner line (1°). The line's orientation was randomly chosen from 360 possible orientations, with the constraint of a different orientation for each stimulus in the sequence. The probe was similar to the preceding three stimuli, and its initial orientation was also determined randomly. A baseline (uncrowded) condition, in which only the target was presented, was also included. The luminance of all stimuli was 0.01 cd/m^2, except for Experiment 3, in which the luminance of the first and third stimuli (i.e., the distractors) was 100 cd/m^2. The background was a uniform gray (23.5 cd/m^2). The chosen presentation side of the stimuli and the SOA were fixed in each trial but varied between trials.

Procedure

Each trial started with the fixation mark that remained throughout the trial. After 1,000 ms, in the crowded condition, the sequence of three stimuli was presented, with the target being second in the stream. In Experiments 1 and 3, the presentation duration of each stimulus was 75 ms, and in Experiment 2 it was 20 ms. In the uncrowded condition, only the target appeared. The probe was presented 500 ms after the offset of the third stimulus in the sequence. The task was to rotate the probe (using the arrow keys) to reproduce the target's orientation. When observers thought that the probe's orientation matches that of the target, they pressed the space bar and the next trial began. Throughout the trial, observers had to fixate their gaze on the central fixation dot, but once the probe appeared, they could move their eyes. Each observer participated in 60 practice trials and completed 600 experimental trials that included 100 trials for each target–distractors SOA condition and 100 trials for the uncrowded condition.

Model fitting

We excluded trials in which a saccade with an amplitude greater than 1° was executed (1.4% of total trials). For each of the remaining trials, we calculated response error as the difference between the target's orientation and the response given by the observer (e.g., if the target's orientation was 60° and the observer rotated the probe to an orientation of 80°, the error measured in this trial was +20°). Response errors ranged from −180 to 180 degrees. We used the MemToolbox (Suchow et al., 2013) to fit the two-misreport mixture model, with all four parameters (g, sd, β_1, β_2), to the error distribution of the crowded condition of each participant. The model was fitted separately for each SOA of the crowded condition. Here is the full model:

$$p(\theta) = (1 - g - \beta_1 - \beta_2)\varphi_\sigma(\theta) + g/2\pi + \beta_1\varphi_\sigma(\theta_1{*}) + \beta_2\varphi_\sigma(\theta_2{*}) \tag{1}$$

where θ is the response error (i.e., the difference between the reported orientation and the target's orientation); g is the height of the uniform distribution (the proportion of random guessing); φ_σ denotes the circular analog of the Gaussian distribution (the Von Mises distribution) with mean equal to zero (zero error) and standard deviation σ *(sd)*; β_1 is the probability of misreporting the orientation of the distractor that preceded the target; and β_2 is the probability of misreporting the orientation of the distractor that followed the target. Finally, $\theta_1{*}$ is the error relative to the orientation of the preceding distractor, and $\theta_2{*}$ is the error relative to the orientation of the succeeding distractor.

Because the uncrowded condition did not include distractors, we fitted the error distribution of this condition with a mixture model that has two free parameters (g, sd):

$$p(\theta) = (1 - g)\varphi_\sigma(\theta) + g/2\pi \tag{2}$$

As can be seen in Fig. 2, with all the three experiments of this study, the model fits the data well, confirming that this continuous-report task and the mixture model were adequate for the study of temporal crowding. To further ensure we were using the optimal model for our data, we compared the two-misreport mixture model (Eq. 1) with the standard misreport mixture model (Eq. 3) that aggregates the contribution of different distractors. The latter model has three free parameters,

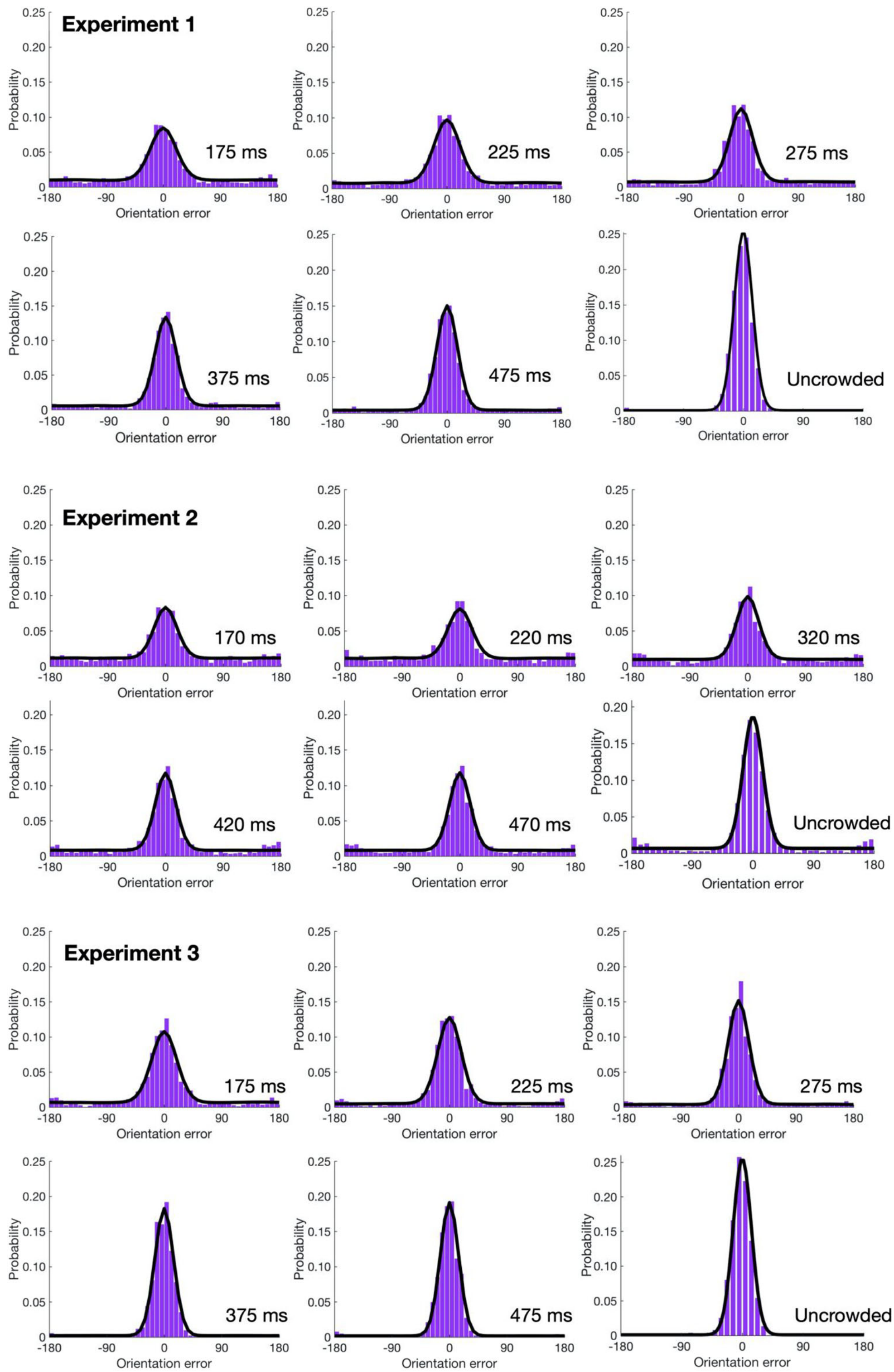

Fig. 2 Mean error distribution and model fits (in black) for the uncrowded condition and each SOA of the crowded condition in Experiments 1–3

the standard deviation *(sd),* the guessing rate (g), and the rate of mistakenly reporting the orientation of one of the distractors (β). Here is the full model, as proposed by Bays et al. (2009)

$$p(\theta) = (1{-}g{-}\beta)\varphi_\sigma(\theta) + g/2\pi + \beta/m\sum_{i}^{m}\varphi_\sigma(\theta_i{}^*) \qquad (3)$$

where m is the number of distractors (two in our study), and $\theta_i{}^*$ is the error relative to the orientation of the i^{th} distractor. The models were compared using the Akaike information criteria with correction (AICc) that includes a penalty term for each additional model parameter.

Experiment 1

The goal of Experiment 1 was to examine which parameter/s of the model will be modulated by temporal crowding when all the stimuli in the sequence have the same luminance (black) and the presentation duration of each stimulus in the sequence is relatively long (75 ms).

Results

First, we compared the fits of the models described above for each participant and each SOA. The two-misreport mixture model outperformed the standard misreport mixture model. A one-tailed paired t test showed that the AICc values of the former were significantly lower than those of the latter, $t(14) = 2.22$, $p = .022$, $d_z = 0.57$. We, therefore, continued with the two-misreport mixture model.

To examine which parameters are affected by temporal crowding, we performed a one-way (SOA) repeated-measures ANOVA on each of the model parameters (blue curves of Fig. 3). A significant effect of SOA was found for the *sd*, $F(4, 56) = 6.05$, $p = .0004$, $\eta_p{}^2 = 0.30$; β_1, $F(4, 56) = 9.39$, $p < .0001$, $\eta_p{}^2 = 0.40$; and β_2, $F(4, 56) = 15.67$, $p < .0001$, $\eta_p{}^2 = 0.53$, parameters, but not the g parameter, $F(4, 56) = 1.97$, $p = .1113$, $\eta_p{}^2 = 0.13$. When SOA was relatively short and the stimuli were crowded in time, *sd*, β_1, and β_2 were high, suggesting that the precision of the target's encoding was low and the substitution rate with both preceding and succeeding distractors was high. As the SOA got longer (i.e., temporal crowding decreased) precision was enhanced and substitution errors were reduced. Still, even with the SOA of 475 ms, *sd* was significantly larger than that observed in the uncrowded condition, $t(14) = 3.62$, $p = .0014$, $dz = 0.936$, extending the impact of temporal crowding into even longer durations than what was found thus far (Tkacz-Domb & Yeshurun, 2017; Yeshurun et al., 2015). In contrast, the lack of a significant SOA effect with the guessing rate implies that temporal crowding may not affect the target's SNR.

Interestingly, the model fitting to the data further suggests that the preceding distractor was mistakenly reported as the target much more often than the succeeding distractor (blue curves in Fig. 3c vs. 3d). A two-way (SOA × distractor) repeated-measures ANOVA performed on the rate of substitution errors confirmed that substitution errors were significantly more prevalent with the preceding than succeeding distractor, $F(1, 14) = 6.33$, $p = .0247$, $_p{}^2 = 0.31$, and this was not qualified by SOA (i.e., no significant interaction with SOA, $F(4, 56) = 1.24$, $p = .3028$, $_p{}^2 = 0.08$.

Experiment 2

In Experiment 2, stimuli duration was shorter. The goal of this experiment was threefold. First, it allowed us to test the effect of stimulus duration on temporal crowding. Second, as will be detailed in the Discussion section, a shorter stimuli duration allows a more straightforward comparison between temporal crowding and ordinary masking. Third, given that the guessing rate found in Experiment 1 was rather low, one might wonder whether the lack of an SOA effect on the guessing rate reflects a floor effect. Shortening display duration should decrease the SNR, which should increase the overall guessing rate, thereby avoiding floor effect as an alternative explanation. Thus, this experiment was similar to Experiment 1 apart from shortening the duration of each orientation stimulus to 20 ms and consequently employing slightly different SOAs.

Results

Trial exclusion followed the same criterion as in Experiment 1 (1.7% of total trials), and we performed the same modeling procedure and statistical analyses. Similar to Experiment 1, a one-tailed paired t test showed that the AICc values of the two-misreport mixture model were significantly lower than those of the standard misreport mixture model, $t(16) = 1.9$, $p = .038$, $dz = 0.467$, and we therefore continued with the former. As can be seen in Fig. 3, the pattern of SOA effects on the model parameters in this experiment (green curves) is very similar to Experiment 1 (blue curves): *sd,* β_1, and β_2 decreased significantly as the SOA increased, $F(4, 64) = 2.56$, $p = .0468$, $_p{}^2 = 0.14$; $F(4, 64) = 5.82$, $p = .0005$, $_p{}^2 = 0.27$; and $F(4, 64) = 21.17$, $p < .0001$, $_p{}^2 = 0.57$, respectively, but there was no significant SOA effect on the guessing rate, $F(4, 64) = 1.72$, $p = 0.1572$, $_p{}^2 = 0.096$, even though the overall guessing rate increased in comparison to Experiment 1. Also similar to Experiment 1, substitution errors were more prevalent with the preceding than succeeding distractor, $F(1, 16) = 18.28$, $p = 0.0006$, $_p{}^2 = 0.53$, regardless of SOA, $F(4, 64) = 1.99$, $p = .1066$, $_p{}^2 = 0.11$. These results suggest that shortening the stimuli duration does not elicit a qualitative change in the way in which temporal crowding affects these parameters. This

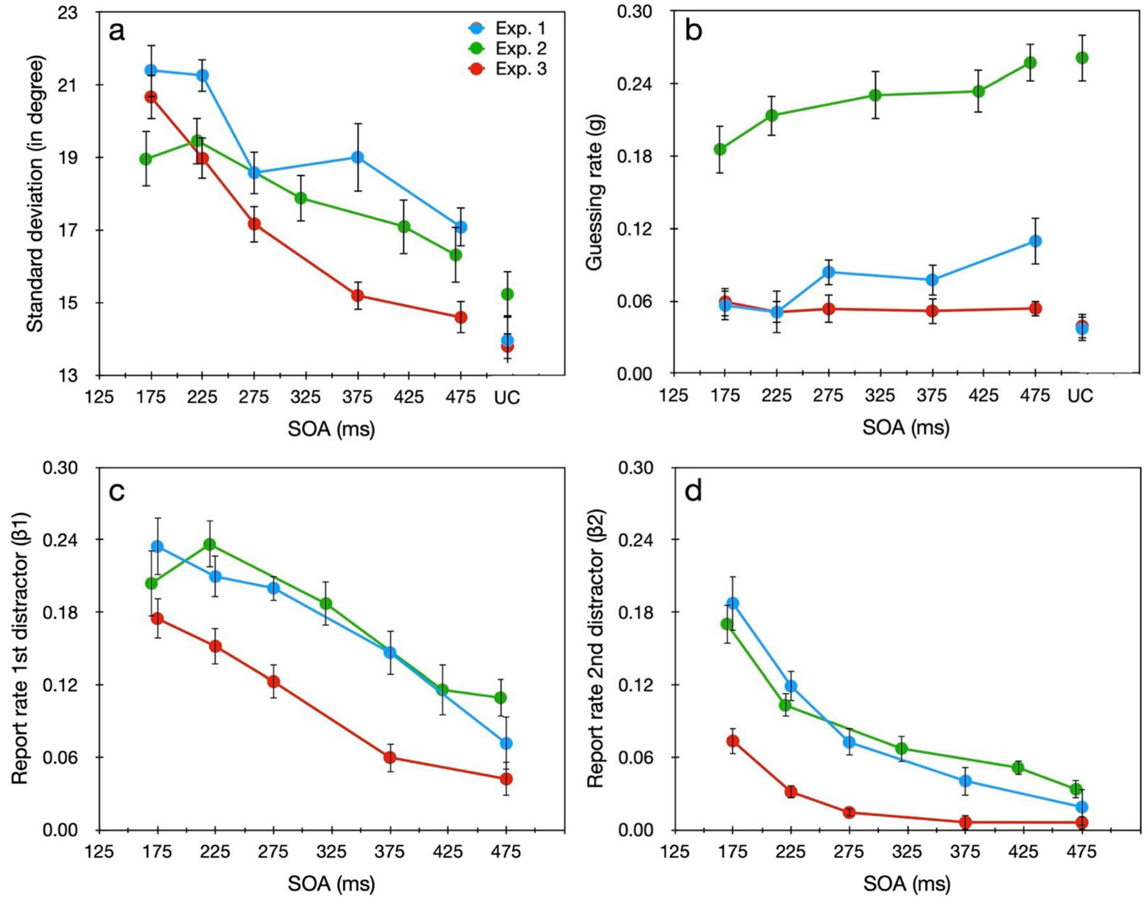

Fig. 3 The averages of the estimated parameters (**a**) *sd* (**b**) *g* (**c**) β_1 (**d**) β_2 as a function of SOA in the crowded condition, and the uncrowded condition (UC) in Experiment 1 (blue), Experiment 2 (green), and Experiment 3 (red). The data points corresponding to the *g* parameter of the uncrowded condition of Experiments 1 and 3 overlap. Error bars correspond to one within-subject standard error (Cousineau, 2005). (Color figure online)

further indicates that the lack of a significant effect of temporal crowding on the SNR, observed in Experiment 1, does not merely reflect a floor effect.

Experiment 3

Previous studies have demonstrated that spatial crowding decreases when the target and flankers are dissimilar (e.g., Chakravarthi & Cavanagh, 2007; Kooi et al., 1994; Levi & Carney, 2009). In this experiment, we examined whether temporal crowding is also reduced when the target and distractors are dissimilar. Thus, this experiment was identical to Experiment 1, except that the target was black and the distractors were white.

Results

Trial exclusion followed the same criterion as in Experiment 1 (1.7% of total trials), and we performed the same modeling procedure and statistical analyses. In this experiment, the AICc values of the two-misreport mixture model were also lower than those of the standard misreport mixture model, but this difference did not reach statistical significance, $t(14) = 0.6$, $p = .3$, $dz = 0.17$, likely due to a floor effect, as with both models the rate of misreporting a distractor was low due to target–distractor dissimilarity (more about this below). Notwithstanding the lack of significant difference, because the two-misreport mixture model provides more information than the standard misreport mixture model, we continued the analyses with this model.

Once again, these analyses revealed a similar pattern of results to that found in Experiment 1 (see Fig. 3 red curves): the *sd*, β_1, and β_2 parameters changed significantly with SOA—$F(4, 56) = 19.796$, $p < .0001$, $\eta_p^2 = 0.59$; $F(4, 56) = 12.703$, $p < .0001$, $\eta_p^2 = 0.48$; $F(4, 56) = 14.397$, $p < .0001$, $\eta_p^2 = 0.51$, respectively—but not the *g* parameter, $F(4, 56) = 0.083$, $p = .987$, $\eta_p^2 = 0.006$. Hence, temporal crowding reduced encoding precision and increased substitution rate, but it did not affect the SNR. Like before, substitution errors were more frequent with the preceding than succeeding distractor, $F(1, 14) = 43.59$, $p < .0001$, $\eta_p^2 = 0.76$, but this time there was also an interaction with SOA, $F(4, 56) = 3.42$, $p = .0143$, $\eta_p^2 = 0.196$, likely due to a floor effect (see Fig. 3c–d, red curves).

To examine directly the effect of target–distractors similarity on temporal crowding we conducted a two-way mixed-design ANOVA on the parameters generated for Experiments 1 and 3, with SOA as a within-subject variable and similarity as a between-subject variable. Not surprisingly given the above analyses, a significant main effect of SOA emerged for the *sd*, β_1, and β_2 parameters—$F(4, 112) = 20.67$, $p < .0001$, $\eta_p^2 = 0.43$; $F(4, 112) = 20.53$, $p < .0001$, $\eta_p^2 = 0.42$; $F(4, 112) = 26.18$, $p < .0001$, $\eta_p^2 = 0.48$, respectively—but not for *g*, $F(4, 112) = 1.33$, $p = .26$, $\eta_p^2 = 0.05$. However, with both *sd* and *g* parameters there was no significant main effect of similarity—$F(1, 28) = 2.62$, $p = .12$, $\eta_p^2 = 0.09$; $F(1, 28) = 0.96$, $p = .34$, $\eta_p^2 = 0.03$, respectively—nor an SOA × Similarity interaction, $F(4, 112) = 1.53$, $p = .198$, $\eta_p^2 = 0.05$; $F(4, 112) = 1.42$, $p = .23$, $\eta_p^2 = 0.05$, respectively. These results suggest that the detrimental effect of temporal crowding on the encoding precision is present regardless of target–distractors similarity, and the same holds for the absence of an effect regarding the SNR. In contrast, with the β_1 and β_2 parameters, the main effect of similarity reached significance, $F(1, 28) = 4.26$, $p = .048$, $\eta_p^2 = 0.13$; $F(1, 28) = 13.42$ $p = .001$, $\eta_p^2 = 0.32$, respectively: Substitution errors were more frequent when the target and the distractors were similar than when they were dissimilar. Additionally, with the β_2 parameter there was also a significant interaction (see Fig. 3d red vs. blue curves), which is likely due to a floor effect, β_1, $F(4, 112) = 0.68$, $p = .6$, $\eta_p^2 = 0.02$; β_2, $F(4, 112) = 4.76$ $p = .0014$, $\eta_p^2 = 0.15$. Hence, a dissimilarity benefit emerged also for temporal crowding, but mainly regarding substitution rate with the succeeding distractor.

Discussion

As was previously demonstrated using a discrete-report task (Tkacz-Domb & Yeshurun, 2017; Yeshurun et al., 2015), identifying the target's orientation was impaired by the presence of preceding and succeeding distractors even when the SOA between them was considerably longer than the limits of ordinary masking. In fact, because the performance was worse in the crowded than uncrowded condition even with an SOA of 475 ms, the current study further extends the temporal range of crowding. Importantly, in this study we were able to go beyond showing mere performance decrement due to temporal crowding. Specifically, we found that temporal crowding degraded the precision of the target's encoding, increased substitution errors with both preceding and succeeding distractors, but did not affect the guessing rate. This pattern of results was similar when the target and distractors had the same luminance and relatively long duration (Experiment 1), when their duration was considerably shorter (Experiment 2), and when the target and distractors had different luminance.

The task and analyses employed here further allowed us to consider the possibility that similar processes underlie temporal crowding and ordinary masking, only on a different time scale. This is because Agaoglu et al. (2015) examined several types of ordinary masking using similar task and analyses, but with much shorter target-mask SOAs (up to 110 ms). Among the masking types examined, pattern making by structure is most relevant for our study because the structure mask shares features with the target. The primary effect of SOA that emerged with structure masking was a significant increase in guessing rate, suggesting that structure masking mainly decreases the SNR. There was no significant effect of SOA on the encoding precision or substitution errors. Thus, Agaoglu et al.'s pattern of results is basically opposite to ours, suggesting that masking and temporal crowding reflect distinct processes.

The task and analyses adopted in this study also allowed us to compare crowding across time and space because Ester et al. (2014) used similar methods, only they examined spatial crowding and therefore the distractors appeared simultaneously with the target at adjacent locations. Thus, whereas we varied target–distractor distance in time, they varied target–distractor distance in space (in their Experiment 3). Unlike our results, they found significant effects of target–distractor distance on the guessing and substitution rates, but not on precision. These different patterns of results suggest that these two types of crowding reflect different processes. However, like us, Ester et al. found in their Experiment 2 that the manipulation of target–distractor similarity mainly affected substitution rate. Thus, temporal and spatial crowding show some commonalities.

The lack of an SOA effect on the guessing rate suggests that temporal crowding may not affect the SNR. This further suggests that temporal crowding, likely, does not involve integrating the target's and the distractors' signals into a single unit, as was suggested for ordinary masking over short SOAs (e.g., Enns, 2004). Instead, the performance decrement brought about by temporal crowding seems to have two components: (1) In all three experiments, shorter SOAs increased the frequency of reporting a distractor instead of the target. This seems akin to substitution masking in which the target representation is replaced with that of the mask (e.g., Di Lollo et al., 2000). However, substitution masking typically refers to backward masking, while we found substitution errors also for the preceding distractor. In fact, more substitution errors were found with the preceding than succeeding distractor. This finding may reflect source confusion in the time domain. Several studies suggested that spatial crowding is at least partially due to confusion regarding the spatial position of the target (e.g., Ester et al., 2014; Strasburger & Malania, 2013). Perhaps temporal crowding is also partially due to source confusion, only here the confusion is about the onset time of each stimulus rather than its spatial position. If so, our data suggest that there is a kind of recency effect: There is less source confusion regarding the most recent item. (2) In all three experiments, shorter SOAs resulted in a significantly larger *sd*. This suggests that the distractors interfere with the processing of the target, before this processing is completed, thereby reducing the precision of the target's encoding. The fact that precision was lower in the crowded than uncrowded condition even with the longest SOA (475 ms) suggests that the processing of the target is not completed even after such a long period. This finding is consistent with other demonstrations of long-lasting temporal interactions (Otto et al., 2009; Scharnowski et al., 2009), and it raises a need to reconsider conclusions that are based on evidence of very fast visual processing. Although previous evidence undoubtedly demonstrated that some representation of visual information may be available after ~150 ms, our findings show that a robust and stable representation requires considerably longer processing time.

Finally, one may wonder whether temporal crowding is related to the attentional blink phenomenon, in which the identification of a second target is impaired when the temporal interval between it and a first target is within the range of about 200–600 ms (e.g., Chun & Potter, 1995; Raymond et al., 1992; for a recent review, see Snir & Yeshurun, 2017). However, the attentional blink phenomenon is fundamentally different from temporal crowding. With the attentional blink, a faster stream of stimuli (SOAs around 100 ms) is presented, and the participants are required to report two targets. The identification impairment is observed for the second target, and it is typically attributed to the need to encode the first target. No impairment is observed for the first target within this range of temporal intervals, even though other nontarget items precede and succeed it. In contrast, with temporal crowding, the participants are required to report only a single target and therefore the observed identification impairment cannot be attributed to the need to encode another target.

In sum, combining a continuous-report task and mixture model analysis with temporal crowding displays allowed us to examine, for the first time, the nature of the impairment brought about by 'pure' temporal crowding. We found that temporal crowding impaired the precision of the target's encoding and increased substitution errors but had no effect on the target's SNR. Moreover, this pattern of results is qualitatively different than that found for ordinary masking and spatial crowding, suggesting that these three phenomena are mediated by different processes. The fact that stimuli separated from the target by almost half a second still affected the quality of the target's representation suggests that the processing of visual information requires a considerably longer time to complete than the current belief.

Author note This research was funded by the Israel Science Foundation (ISF) grant to Y.Y. (No. 1780/19).

S.T.-D. designed and conducted the experiments, analyzed and interpreted the data, and wrote the manuscript. Y.Y. designed the experiments, analyzed and interpreted the data, and wrote the manuscript.

References

Agaoglu, S., Agaoglu, M. N., Breitmeyer, B., & Ogmen, H. (2015). A statistical perspective to visual masking. *Vision Research, 115,* 23–39. https://doi.org/10.1016/j.visres.2015.07.003

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. P*sychological Science*, *28*(11), 1547–1562. https://doi.org/10.1177/0956797617723724

Asplund, C. L., Fougnie, D., Zughni, S., Martin, J. W., & Marois, R. (2014). The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychological science, 25*(3), 824–831. https://doi.org/10.1177/0956797613513810

Bankson, B. B., Hebart, M. N., Groen, I. I., & Baker, C. I. (2018). The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks. *NeuroImage, 178,* 172–182. https://doi.org/10.1016/j.neuroimage.2018.05.037

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(2)7, 1–11. https://doi.org/10.1167/9.10.7

Bonneh, Y. S., Sagi, D., & Polat, U. (2007). Spatial and temporal crowding in amblyopia. *Vision Research, 47*(14), 1950–1962. https://doi.org/10.1016/j.visres.2007.02.015

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178. https://doi.org/10.1038/226177a0

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Breitmeyer, B. (1984). *Visual masking: An integrative approach*. Oxford University Press.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, *62*, 1572–1595. https://doi.org/10.3758/BF03212157

Breitmeyer, B. G., & Ogmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision*. Oxford University Press.

Chakravarthi, R., & Cavanagh, P. (2007). Temporal properties of the polarity advantage effect in crowding. *Journal of Vision, 7*(2):11, 1–13. https://doi.org/10.1167/7.2.11

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 109–127. https://doi.org/10.1037/0096-1523.21.1.109

Cousineau, D. (2005). Confidence intervals in within- subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*, 481–507. https://doi.org/10.1037/0096-3445.129.4.481

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research*, *44*(12), 1321–1331. https://doi.org/10.1016/j.visres.2003.10.024

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*(9)*,* 345–352. https://doi.org/10.1016/S1364-6613(00)01520-5

Ester, E. F., Klee, D., & Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 1022–1033. https://doi.org/10.1037/a0035377

Gorea, A. (1987). Masking efficiency as a function of stimulus onset asynchrony for spatial-frequency detection and identification. *Spatial Vision, 2*(1), 51–60. https://doi.org/10.1163/156856887X00051

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20*(4), 464–472. https://doi.org/10.1111/j.1467-9280.2009.02316.x

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310*(5749), 863–866. https://doi.org/10.1126/science.1117593

Kleiner, M., Barnard, D. H., & Pelli D. G. (2007). What's new in Psychtoolbox-3? *Perception, 36*(ECVP Abstract Supplement), 1–16.

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial vision, 8*(2), 255–279. https://doi.org/10.1163/156856894X00350

Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: Why bigger is better. *Current Biology, 19*(23), 1988–1993. https://doi.org/10.1016/j.cub.2009.09.056

Otto, T. U., Ögmen, H., & Herzog, M. H. (2009). Feature integration across space, time, and orientation. *Journal of Experimental Psychology: Human Perception and Performance, 35*(6), 1670–1686. https://doi.org/10.1037/a0015798

Rashal, E., & Yeshurun, Y. (2014). Contrast dissimilarity effects on crowding is not simply another case of target saliency. *Journal of Vision, 14*(6):9, 1–12.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849–860. https://doi.org/10.1037/0096-1523.18.3.849

Scharnowski, F., Rüter, J., Jolij, J., Hermens, F., Kammer, T., & Herzog, M. H. (2009). Long-lasting modulation of feature integration by transcranial magnetic stimulation. *Journal of Vision, 9*(6):1, 1–10. https://doi.org/10.1167/9.6.1

Shechter, A., & Yashar, A. (2021). Mixture model investigation of the inner-outer asymmetry in visual crowding reveals a heavier weight towards the visual periphery. *Scientific Reports, 11,* Article 2021. https://doi.org/10.1038/s41598-021-81533-9

Snir, G., & Yeshurun, Y. (2017). Perceptual episodes, temporal attention, and the role of cognitive control: Lessons from the attentional blink. In C. Howard (Ed.), *Progress in brain research: Vol. 236. Temporal sampling and representation updating* (pp. 53–74). Academic Press. https://doi.org/10.1016/bs.pbr.2017.07.008

Strasburger, H. (2005). Unfocussed spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, *5*(11):8, 1024–1037. https://doi.org/10.1167/5.11.8

Strasburger, H., & Malania, M. (2013). Source confusion is a major cause of crowding. *Journal of Vision, 13*(1):24, 1–20. https://doi.org/10.1167/13.1.24

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision, 13*(10):9, 1–8. https://doi.org/10.1167/13.10.9

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature 381,* 520–522. https://doi.org/10.1038/381520a0

Tkacz-Domb, S., & Yeshurun, Y. (2017). Spatial attention alleviates temporal crowding, but neither temporal nor spatial uncertainty are necessary for the emergence of temporal crowding. *Journal of Vision, 17*(3):9, 1–12. https://doi.org/10.1167/17.3.9

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Science, 15*(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005

Yeshurun, Y., & Rashal, E. (2010). Precueing attention to the target location diminishes crowding and reduces the critical distance. *Journal of Vision, 10*(10):16, 1–12. https://doi.org/10.1167/10.10.16

Yeshurun, Y., Rashal, E., & Tkacz-Domb, S. (2015). Temporal crowding and its interplay with spatial crowding. *Journal of Vision, 15*(3):11, 1–16. https://doi.org/10.1167/15.3.11

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235. https://doi.org/10.1038/nature06860

Open practices statement The data that support the findings of this study are available in the Open Science Framework repository (https://osf.io/svbtx/?view_only=70859085fce04334a2b38d9e2ec85d30). The experiments reported in this article were not preregistered.