Cognition 268 (2026) 106374

Contents lists available at ScienceDirect

Cognition

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

Temporal dynamics of integration and individuation: Insights from temporal averaging and crowding

Ilanit Hochmitz [a,*], Yaffa Yeshurun [a,b], Amit Yashar [c,d]

[a] The Institute of Information Processing and Decision Making, University of Haifa, Israel
[b] School of Psychological Sciences, University of Haifa, Israel
[c] The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Israel
[d] Department of Special Education, University of Haifa, Israel

ARTICLE INFO

Keywords:
Vision
Temporal crowding
Individuation
Pooling
Averaging
Masking

ABSTRACT

Individuating a single item presented within a continuous sequence of items requires segregating its signal from that of the other items. In contrast, representing a global aspect of the sequence, such as its average orientation, involves integration of information across time. Individuation and integration allow us to focus on individual events while maintaining an overall perception of our environment. To examine the relations between temporal averaging and individuation, we measured orientation averaging over short and long timescales using the same stimuli and orientation-estimation procedure previously used to measure individuation. Participants reported the average orientation of a sequence of three oriented items separated by either short (SOAs<150 ms) or long intervals (SOAs>150 ms). Analysis of the error distribution and mixture-modeling revealed distinct patterns of results for the different tasks and timescales, but also some similarities, particularly for the short timescale. In this timescale, the relative contribution of each individual item to the final response was similar across tasks, indicating the involvement of low-level factors operating regardless of the task. With the long timescale, the two tasks showed dissociable pattern across all performance aspects, except guessing rate, indicating that long-scale individuation and averaging engage mainly higher-level, task-related processes. Importantly, regardless of timescale, estimation errors in these tasks were best described by different models: in integration they primarily reflected unequal weighting of the averaged items, whereas in individuation they reflected imprecise target encoding with occasional misreports of distractors. Together, the findings reveal dissociable dynamics for integration and individuation.

1. Introduction

At every moment, our visual system confronts a deluge of information from our environment, which must be processed efficiently to support our goals and actions. Often, we are interested in only a subset of this information, for instance, when estimating the speed of one car among many on a busy road. In these cases, the visual system must individuate objects in time and space to focus on specific elements. In other situations, however, we aim to integrate or pool some information across time and space, gathering details to form a cohesive understanding of our surroundings, like estimating the average speed of all cars on the road. Understanding how the visual system navigates between individuation and integration is crucial for understanding how the visual system gathers information on specific objects and features while maintaining a cohesive image of the environment.

While individuation and integration occur both in space and time, the majority of studies have investigated these functions when they occur in the spatial domain – when items appear simultaneously at adjacent spatial locations. In the spatial domain, individuation and integration, particularly in the sense of averaging, are often tested through the phenomena of spatial crowding and ensemble perception, respectively. Spatial crowding refers to the difficulty in identifying a feature or an object because of its proximity to other features and objects (e.g., Chung, 2016; Ester et al., 2014a, 2014b; Greenwood & Parsons, 2020; Kewan-Khalayly et al., 2022; Lev & Polat, 2015; Rashal & Yeshurun, 2014; Shechter et al., 2024; Yeshurun & Rashal, 2010; reviewed by: Pelli, 2008; Strasburger, 2020; Whitney & Levi, 2011). Ensemble perception often refers to the ability of participants to extract summary

* Corresponding author at: The Institute of Information Processing and Decision Making, University of Haifa, Haifa 3498838, Israel.
E-mail address: ilanit57@gmail.com (I. Hochmitz).

https://doi.org/10.1016/j.cognition.2025.106374
Received 6 February 2025; Received in revised form 26 October 2025; Accepted 8 November 2025
Available online 25 November 2025

statistical information of a group of items (reviewed by Whitney & Yamanashi Leib, 2018). Most relevant for the current study is the ability to accurately estimate the average value of an ensemble of items, and in the spatial domain it was demonstrated with various features such as size, orientation, color, and position (e.g., Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003; de Gardelle & Summerfield, 2011; Greenwood et al., 2009; Robitaille & Harris, 2011).

Interestingly, averaging across space, particularly when occurring passively and obligatorily in the visual periphery, has been proposed to explain observers' failures to individuate the target feature, such as in the case of spatial crowding. These views assert that crowding results from an early, compulsory pooling or averaging of information across the target and flankers, such that higher-level processing can only access the averaged information. Such views predict averaging errors in crowding, i.e., reporting the average value of presented features instead of the target value (e.g., Banno & Saiki, 2012; Lin et al., 2022; Parkes et al., 2001). However, other spatial crowding models propose more elaborate processes of information pooling that do not simply predict feature averaging errors (e.g., Rosenholtz, 2016). Furthermore, recent studies have shown that errors in crowding are contingent on feature type. Specifically, while orientation and color crowding often lead to substitution (misreport) errors—misreporting a flanker feature as the target—spatial frequency crowding results in averaging errors (Yashar et al., 2019; Yashar & Carrasco, 2025).

Pooling, particularly averaging, and individuation were also observed in the temporal domain, i.e., when stimuli are presented sequentially to the same location within a given trial rather than simultaneously to different locations. Nevertheless, no study has directly examined the link between temporal averaging—i.e., within-trial sequential mean estimation—and individuation across a wide range of inter-item intervals (henceforth: timescales), using the same stimuli and trial procedure. This is particularly so when considering 'pure' temporal interactions—when only a single item is presented at a given moment. With respect to temporal averaging, participants are able to estimate quite precisely the average size (Corbett & Oriet, 2011; Hubert-Wallander & Boynton, 2015; Khayat et al., 2023; Khayat & Hochstein, 2018), orientation (Khayat & Hochstein, 2018), length (Tong et al., 2019; Tong & Dubé, 2022a, 2022b), location and facial expressions (Hubert-Wallander & Boynton, 2015) of items presented in a sequence, suggesting that information can be integrated and pooled over time. Investigations of temporal averaging often rely on a single stimulus-onset-asynchrony (SOA). This SOA could be either short (i.e., ~100 ms; e.g., Corbett & Oriet, 2011), intermediate (i.e., ~200 ms; e.g., Khayat & Hochstein, 2018; Navajas et al., 2017; Wyart et al., 2012; Khayat et al., 2023) or long (i.e., 600 ms or 1500 ms; e.g., Do et al., 2022; Tong & Dubé, 2022a, 2022b), but no study systematically manipulated the SOA. As for temporal individuation, the main challenge is the interference generated by the preceding and/or succeeding items, and the characteristics of this interference depend on the temporal scale. Specifically, short-scale interference, up to an SOA of 100-150 ms, commonly known as visual masking, is associated with the reduction or even elimination of the task-relevant target perception (e.g., Breitmeyer, 1984; Breitmeyer & Öğmen, 2000, 2006; Enns, 2004; Enns & Di Lollo, 2000; Gorea, 1987). Despite the large body of work around it, the underlying mechanisms of masking are still not fully understood. More recently, it has been shown that participants may fail to identify the target even with SOAs that are longer than 150 ms (e.g., Bonneh et al., 2007; Sahar & Yeshurun, 2024; Yeshurun et al., 2015). This long-lasting interference, known as temporal crowding, was found even with an SOA of 475 ms, far beyond a typical masking effect (Hochmitz et al., 2024; Tkacz-Domb & Yeshurun, 2021). Temporal crowding can occur when there is no temporal uncertainty, or when attention is allocated to the target location (Tkacz-Domb & Yeshurun, 2017).

Do masking and temporal crowding indeed represent different types of interference, or is temporal crowding merely a manifestation of prolonged masking? To address this question, Hochmitz et al. (2024) directly compared masking and temporal crowding using the same task and stimuli with only the range of SOAs differentiating the two types of interference. The task involved estimating the orientation of a target defined as the middle item in a sequence of three stimuli, all presented to the same location within a given trial. The SOAs were either long (Tkacz-Domb & Yeshurun, 2021; 170–470 ms) or short (Hochmitz et al., 2024; 40–120 ms). Mixture modeling analysis (Shechter & Yashar, 2021; see Method section for details) revealed qualitatively different patterns of interference for the short and long temporal scales. Interference over long intervals mainly resulted in degraded precision of the target encoding into memory while interference over short intervals mainly affected the guessing rate. Both instances of interference modulated substitution errors (reporting a wrong item), but in a markedly disparate manner. With long-scale interference, substitution errors decreased with increasing SOAs for both non-target items. In contrast, with short-scale interference, substitution errors increased with SOA for the non-target item that preceded the target, but decreased for the one that succeeded the target. Moreover, a differential effect of items' similarity emerged for temporal crowding and masking (i.e., when the contrast polarity of the target and non-targets was different). With the former, only the encoding precision was affected, but with the latter only the guessing rate was affected (Hochmitz et al., 2024). Finally, while the magnitude of masking is considerably reduced at the center of the visual field in comparison to the periphery (e.g., Breitmeyer & Ganz, 1976; Coates et al., 2018; Francis, 2003; Matthews, 1973; Strasburger, 2020), the magnitude of temporal crowding is similar with central and peripheral presentations (Sahar & Yeshurun, 2024). These various findings support the notion that temporal crowding and visual masking are two distinct phenomena operating through different perceptual processes.

But what about temporal averaging? The current literature does not clearly differentiate the perceptual mechanisms underlying short-scale versus long-scale temporal averaging. While studies of individuation and averaging often use different methodologies, short-scale and long-scale averaging are also typically measured with distinct stimuli, features, and tasks (e.g., Corbett & Oriet, 2011; Hubert-Wallander & Boynton, 2015; Khayat & Hochstein, 2018). Notably, while temporal averaging, within a given trial, has been examined at various timescales including longer intervals, systematic manipulation of SOA to understand how averaging varies across different temporal spans remains largely unexplored.

In the current study, we used the same stimuli, orientation estimation procedure, and mixture modeling approach that was previously used to investigate short- and long-scale interference on stimulus individuation (i.e., masking and temporal crowding; Hochmitz et al., 2024), but this time focusing on orientation averaging across different timescales. Participants were presented with a sequence of three randomly oriented stimuli. They then had to rotate a probe to report the average orientation of the three stimuli. In Experiment 1, the SOAs ranged from 40 ms to 120 ms, examining averaging within the timescale of visual masking. In Experiment 2, the SOAs ranged from 180 ms to 430 ms examining averaging within the timescale of temporal crowding. This experimental design allowed us to determine whether or not short- and long-scale averaging yield comparable patterns. If the same combination of processes mediates temporal averaging regardless of the temporal scale, we expect to find a similar pattern of SOA effects for the different aspects of performance established with mixture modeling (e.g., guessing rate, precision). In contrast, if the combination of the processes involved in temporal averaging changes with the timescale, then the pattern of SOA effects may be different for short and long scales. This experimental design also allowed us to explore the relationship between averaging and individuation within each temporal scale. For instance, if averaging underlies temporal crowding, as was suggested in the spatial domain, we expect to observe SOA effects for both long-scale averaging and temporal crowding—a significant SOA modification of precision and a similar SOA effect on the relative contribution of each item in the sequence to the reported estimation. A different pattern of SOA effects

for individuation and averaging would suggest that these phenomena reflect a different combination of processes. Furthermore, we fit the same weighted average model to the data of both tasks. If this model is a good account for the averaging data and averaging mediates individuation, this model should fit similarly well the individuation data.

2. Methods

2.1. Participants

Twenty-two students (15 females, 7 males; age range: 19–37; mean age: 24.8 years) participated in Experiment 1, and twenty participated in Experiment 2 (15 females, 5 males; age range: 19–37; mean age: 23 years). All participants provided their informed consent and received course credit or monetary payment for their participation. The participants were naïve as to the purpose of the experiment and reported normal or corrected-to-normal vision. The sample size choice was based on a power analysis conducted with the R *pwr* package (Champely, 2020). This analysis indicated that 16 participants is the minimum sample size required for the examination of temporal averaging effects with a power of 0.95 and $\alpha = 0.05$. The F values, degrees of freedom, and effect sizes used in this analysis were based on Khayat and Hochstein (2018; $F(1,38) = 18.93$, $\eta_p^2 = 0.33$, $N = 39$). This analysis confirmed that the current study sample size had sufficient statistical power. This study adhered to the Declaration of Helsinki and was approved by the institutional ethics committee (287/19).

2.2. Stimuli and apparatus

Stimuli were presented on a 19″ monitor of an IBM-compatible PC (1024 × 768 resolution at a refresh rate of 85 Hz), using MATLAB and the Psychophysics Toolbox extensions (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli, 1997). Eye movements were tracked from the right eye using an EyeLink 1000 eye tracker (temporal resolution of 1000 Hz; SR Research, Ottawa, Ontario, Canada). A sequence of three stimuli, followed by a probe, was randomly presented to the right or the left of a central fixation circle (diameter 0.3°) at an eccentricity of 9° (Fig. 1). Each stimulus comprised a black circle (0.01 cd/m^2; diameter: 2°) with an oriented inner line (1°). The orientation of the line was randomly chosen from 360 possible orientations while maintaining the constraint of a different orientation for each stimulus in the sequence. Stimuli were separated by an SOA. Possible SOAs: 40, 60, 80,100, 120 ms (Experiment 1); 130, 180, 230, 330, 430 ms (Experiment 2). We note that the inclusion of the 130 ms SOA in the long-scale experiment was due to a programming mistake, and therefore these trials were excluded from main statistical analyses.

The probe also featured a black circle with an inner oriented line, with its initial orientation determined randomly and independently. Although the probe's starting value could potentially influence participants' estimates, Tong et al. (2019) did not find an effect of the probe's initial value on line length averaging, and in any case, because this initial value is random and the time between the last stimulus in the sequence and the probe is the same for all SOAs, it cannot account for any systematic SOA effects.

To allow a comparison with the results obtained for short/long scale individuation (i.e., where participants have to segregate the sequence into individual items and report the middle one) this study followed the exact same method used in those studies (Hochmitz et al., 2024; Tkacz-Domb & Yeshurun, 2021) with the only difference being the task. Thus, we included baseline trials in which only a single item appeared. In these baseline trials participants had to report the orientation of the single stimulus. All stimuli were displayed against a uniform gray background (23.5 cd/m^2).

2.3. Procedure

Each trial began with a fixation mark presented for the entire trial. After 1000 ms, the three-stimuli sequence was presented. Each stimulus was presented for 30 ms. The SOA between the sequence items was the same within a trial but varied randomly across trials. The probe was presented 500 ms following the offset of the third stimulus in the sequence. The task was to report the average orientation by clicking with the mouse on the probe's circle outline. Participants could adjust their response as often as required with no time limit. Once they were content with their response, they moved to the next trial by pressing the space bar. The participants had to maintain central fixation until probe onset, during probe presentation, they could move their eyes. Target–distractor SOA was counterbalanced and presented randomly. Altogether, there were 600 experimental trials per experiment, consisting of 100 trials for each SOA condition and additional 100 baseline trials. Additionally, the experimental trials were preceded by a practice session comprising 60 trials that were identical to those in the experimental phase.

2.4. Analysis and models

2.4.1. Data analysis and preprocessing

We calculated the estimation error in each trial by, first, calculating the true value of the circular mean (true average) of the three orientations in each trial. Next, we subtracted the true average value from the estimated value (*error* = *estimated average* – *true average*). Then, for each observer in each condition, we assessed overall performance (Agaoglu et al., 2015; Hochmitz et al., 2024; Sahar & Yeshurun, 2024).

$$\text{Overall performance} = 1 - (\textit{mean absolute error}/180) \quad (1)$$

When the participant produces no errors, the mean absolute error will be zero, resulting in an overall performance value of 1. An overall performance score close to 0.5 suggests that the participant consistently guessed the average orientation, as the mean absolute error would be approximately 90°. Participants whose overall performance score (Eq. 1) was below 0.55 were excluded from further analysis.

2.4.2. Models

We analyzed the error distributions by fitting individual probabilistic-mixture models developed from the standard and standard-with-misreporting models (Bays et al., 2009; Zhang & Luck, 2008). We also tested a new probabilistic model that explains error distribution as a weighted average across the different display items.

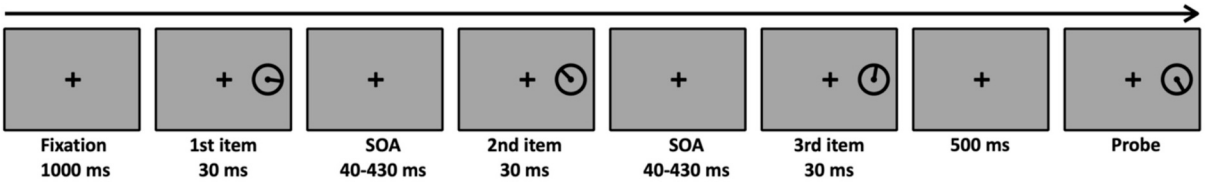

Fig. 1. An example of a single trial in Experiments 1 and 2. In Experiment 1 there were five possible SOAs (40, 60, 80, 100, 120 ms). In Experiment 2 there were also five possible SOAs (130, 180, 230, 330, 430 ms), but the SOA of 130 ms was excluded from the main analyses (see text). The SOA was fixed within a trial but varied between trials. The task was to report the average orientation by rotating the probe line.

2.4.2.1. Standard mixture model (S). The Standard mixture model (Eq. 2) uses a von Mises (circular) distribution to describe the probability density of the estimation around the true-average value and a uniform component to account for guessing. In this model, the probability of reporting an orientation value $P(\widehat{\theta})$ is given by:

$$P(\widehat{\theta}) = (1-\gamma)\phi_{\sigma}(\widehat{\theta}-\theta) + \gamma\left(\frac{1}{n}\right) \quad (2)$$

where $\widehat{\theta}$ is the reported value, θ is the true value (i.e., the true circular average), γ is the proportion of trials in which observers are randomly guessing (reflecting guessing rate), with $n = 360$. ϕ_{σ} is the von Mises distribution with a mean of 0 and a standard deviation (σ) reflecting the precision of initial encoding into memory. The concentration parameter of the von Mises distribution, κ, is converted to σ for interpretability. All orientation differences (e.g., $(\widehat{\theta}-\theta)$) were computed as signed circular differences in the range [−180°, 180°], to account for the periodic nature of orientation space. The model has two free parameters (γ, σ).

2.4.2.2. Weighted Average model (WA). Importantly, we fitted a novel Weighted average model. In the basic version of the Weighted average model, the probability of reporting an orientation value is:

$$p(\widehat{\theta}) = \phi_{\sigma,\overline{\mu}}(\widehat{\theta}-\theta) \quad (3)$$

Where $\phi_{\sigma,\overline{\mu}}$ is a von Mises distribution, with a σ and a mean $\overline{\mu}$, defined as a weighted average:

$$\overline{\mu} = \beta_1\varphi_1 + \beta_2\varphi_2 + (1-\beta_1-\beta_2)\varphi_3 \quad (4)$$

where β_1 and β_2 are weights for the 1st and 2nd item respectively. φ_1, φ_2 and φ_3 are the signed circular differences between the orientation of each item and the circular average of the display θ. This weighted average was computed in circular space, using the standard vector-based method (Fisher, 1995), ensuring that the mean reflects circular topology. Thus, $\overline{\mu}$ reflects the directional bias (in circular space) of the weighted average from the true circular mean on each trial. The model has three free parameters $(\sigma, \beta_1, \beta_2)$.

2.4.2.3. Weighted Average with Guessing model (WU). Furthermore, we fitted a novel Weighted average with guessing model. This model is similar to the previous one, but it assumes that on some of the trials the participants might be guessing. In this model, the probability of reporting an orientation value is:

$$p(\widehat{\theta}) = (1-\gamma)\phi_{\sigma,\overline{\mu}}(\widehat{\theta}-\theta) + \gamma\left(\frac{1}{n}\right) \quad (5)$$

Where $\phi_{\sigma,\overline{\mu}}$ is a von Mises distribution, with a σ and a mean $\overline{\mu}$, defined as a weighted average (Eq. 4). The model has four free parameters $(\sigma, \beta_1, \beta_2, \gamma)$.

2.4.2.4. Weighted Average with Variable Encoding Precision model (WV). We also fitted a novel Weighted average with variable encoding precision model. It assumes that the items in the sequence might be encoded with different precision. In this model, the probability of reporting an orientation value is:

$$p(\widehat{\theta}) = (1-\gamma)\overline{\phi}(\widehat{\theta}-\theta) + \gamma\left(\frac{1}{n}\right) \quad (6)$$

Where ϕ is a von Mises distribution defined as a weighted average of three von Mises distributions:

$$\overline{\phi} = \beta_1\phi_{\sigma_1,\varphi_1} + \beta_2\phi_{\sigma_2,\varphi_2} + (1-\beta_1-\beta_2)\phi_{\sigma_3,\varphi_3} \quad (7)$$

Here σ_1, σ_2, and σ_3 are the encoding precision, and φ_1, φ_2 and φ_3 are the signed circular differences between each item and the true circular average. Each component $\phi_{\sigma_i,\varphi_i}$ is a von Mises distribution in circular space, and the mixture $\overline{\phi}$ preserves the circular topology through proper normalization. The model has six free parameters $(\beta_1, \beta_2, \sigma_1, \sigma_2, \sigma_3, \gamma)$.

2.4.3. Individuation task

To examine whether averaging and individuation across time rely on similar processes we compared the two phenomena within each timescale. With the short temporal scale, we compared the data obtained in Experiment 1 with that of Hochmitz et al. (2024); Experiment 1), and with the long temporal scale we compared the data obtained in Experiment 2 with that of Tkacz-Domb and Yeshurun (2021; Experiment 2). These experiments used similar stimuli, procedures, and SOAs only that participants were required to perform an 'individuation' task—report the orientation of the middle item (i.e., the target) rather than the average orientation. We compare the two tasks with regard to both the overall estimation error (overall performance, Eq. 1) of each task as well as the models' parameters. We fitted the Weighted average with guessing model (WU) to the data of the individuation task and tested how well it can account for data in comparison with the Two-misreport model (Eq. 8) that was originally applied to the individuation task data.

2.4.3.1. Two-misreport mixture model (individuation task data). In the Two-misreport model, the probability of reporting an orientation value is:

$$P(\widehat{\theta}) = (1-\gamma-\beta_1-\beta_2)\phi_{\sigma}(\widehat{\theta}-\theta) + \gamma\left(\frac{1}{n}\right) + \beta_1\phi_{\sigma}(\widehat{\theta}-\varphi_1) + \beta_2\phi_{\sigma}(\widehat{\theta}-\varphi_2) \quad (8)$$

where β_1 and β_2 are the rate of misreporting the orientation of the 1st and 2nd distractors, respectively (the 1st and 3rd items in the sequence) as the orientation of the target, and $(1-\gamma-\beta_1-\beta_2)$ is the rate of reporting the target (the 2nd item). φ_1, φ_2 are the true value of the 1st and 2nd distractors respectively.

Although the Two-misreport and the Weighted average with guessing model describe different operations, they are comparable. Both models have 4 free parameters and both models include report σ (reflecting the encoding precision) and guessing rate (γ). Additionally, both models include parameters that reflect the contribution of each stimulus to the final report, either in the form of the averaging weight assigned for each item (β_1, β_2, 1- β_1 - β_2; Eq. 4) or the rate at which each item was reported (β_1, target report rate = 1– γ–β_1–β_2, β_2; Eq. 8). In all analyses, we annotate these weights in both models according to their temporal order as *S1*, *S2*, and *S3*.

2.4.4. Model-fitting procedure

We used the MemToolbox (Suchow et al., 2013; https://doi.org/10.1167/13.10.9) for model fitting and comparison. To compare models, we calculated the Akaike information criterion with correction (AICc) for the individual fits.

3. Results

Three participants were excluded from Experiment 1 due to a too-low overall performance score (Eq. 1), and three were excluded from Experiment 2; the final number of participants included in the statistical analysis was 19 for Experiment 1 and 17 for Experiment 2. Additionally, 3.6 % of the trials, in Experiment 1, were removed from further analysis due to eye movements (saccade amplitude >1°) or instant pressing of the space bar without clicking on the probe. In Experiment 2, 4.09 % of the trials were excluded for the same reasons. All the experimental trials in the SOA 130 ms condition of Experiment 2 were removed from the main analysis as they do not fall within the range of long-scale processes (i.e., SOA > 150 ms). Thus, for consistency, we will refer to the long-scale range as spanning 180-430 ms. However, the **130 ms condition was not discarded entirely**: it was **included** in the piecewise regression analysis (Section **3.2.3**), which jointly examines short- and long-

scale dynamics across the full SOA range. Plots of parameter estimates across the full SOA range, including the130ms condition, can be found in the Supplementary Materials (**Supplementary Figs. 1–2**). Finally, 'baseline' trials in which only a single item appeared were disregarded; these trials were included in the experiments only to closely follow the procedure of the 'temporal interference' studies (Hochmitz et al., 2024; Tkacz-Domb & Yeshurun, 2021). Mean error distributions for the full SOA range, including baseline and the SOA 130 ms condition, can be seen in Fig. 4. The full data set is available at https://osf.io/zkfjq/?view_only=61fbebbc73474b499d11b6aa6253e480.

3.1. Estimation errors

We start with analyzing the effect of SOA on the overall performance measure (Eq. 1), which represents the participants' overall averaging ability in the current task. A one-way repeated measures ANOVA revealed a significant SOA effect on overall performance in short-scale averaging (Experiment 1; $F(4,72) = 13.451$, $p < .001$, $\eta p^2 = 0.43$). As the SOA increased, overall performance decreased (Fig. 2A, red curve left panel), indicating that participants' ability to report the average orientation is weakened with longer inter-item intervals. In contrast, for long-scale averaging, there was no significant effect of SOA on overall performance ($F(3,48) = 0.713$, $p = .549$, $\eta p^2 = 0.043$; Fig. 2A, red curve right panel). These different patterns of results already suggest scale-dependent differences in how averaging operates. Still, the following analyses of the modeling outcomes provide a more thorough examination of the differences and similarities across temporal scales.

3.2. Probabilistic-mixture models

3.2.1. Model comparisons

Fig. 3 depicts the mean ΔAICc for each model in Experiment 1 (Fig. 3A) and Experiment 2 (Fig. 3B). ΔAICc was calculated as ΔAICc = AICc – min(AICc), where min(AICc) is the AICc of the best-fitting model. In both experiments, the best-fitting model was the Weighted average with guessing model (WU), with significantly lower AICc compared to the second-best performing model, i.e., the Weighted average with no-guessing model (WA; Experiment 1: $t(18) = 3.03$, $p = .007$, Experiment 2: $t(16) = 3.52$, $p = .003$). As can be seen in Fig. 4, the model fits the data well, thus, we continue with analyzing the fitted parameters of the Weighted average with guessing model in both experiments. Fitted parameters across SOAs for the Weighted Average without a guessing model component are presented in the Supplementary Materials (**Supplementary Fig. 3**).

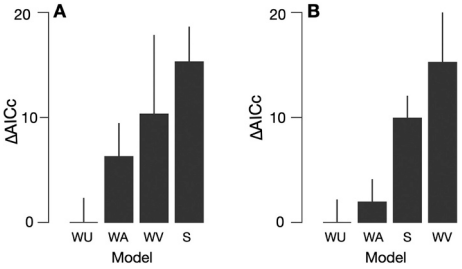

Fig. 3. Model comparisons. Mean AICc for each model subtracted by the AICc of the best fitted model in (**A**) short-scale SOAs (Experiment 1) and (**B**) long-scale SOAs (Experiment 2). ΔAICc = AICc – min(AICc), where min(AICc) is the AICc of the best-fitting model: the Weighted average with guessing model (WU; Eqs. 4, 5). WA = Weighted average with no guessing component model (Eqs. 3, 4); WV = Weighted average with variable precision model (Eqs. 6, 7), S = Standard mixture model (Eq. 2). Error bars represent one standard error.

Additionally, we fitted the Weighted average with guessing model to the individuation task for comparison with the Two-misreport model. In both experiments, the Two-misreport model significantly outperformed the Weighted average with guessing model (short SOA: $t(15) = 7.62$, $p < .001$; long SOA: $t(16) = 10.02$, $p < .001$; see Supplementary Materials, **Supplementary Fig. 4**, for visualization).

3.2.2. Analysis of fitted parameters as a function of SOA

For each experiment, we performed a one-way (SOA) repeated measures ANOVA on each of the Weighted average with guessing model parameters.

3.2.2.1. Standard deviation (σ). A significant effect of SOA on the σ parameter (Fig. 2B, red curve) was found for both short-scale (left panel) and long-scale (right panel) averaging (Experiment 1: $F(4,72) = 3.465$, $p = .012$, $\eta_p^2 = 0.16$; Experiment 2: $F(3,48) = 3.11$, $p = .035$, $\eta_p^2 = 0.16$), but of a very different nature. With short-scale averaging the σ increased as the SOA increased, suggesting that the encoding precision of the average was degraded with increasing inter-item intervals. With the long-scale, however, this effect showed no clear pattern, making it

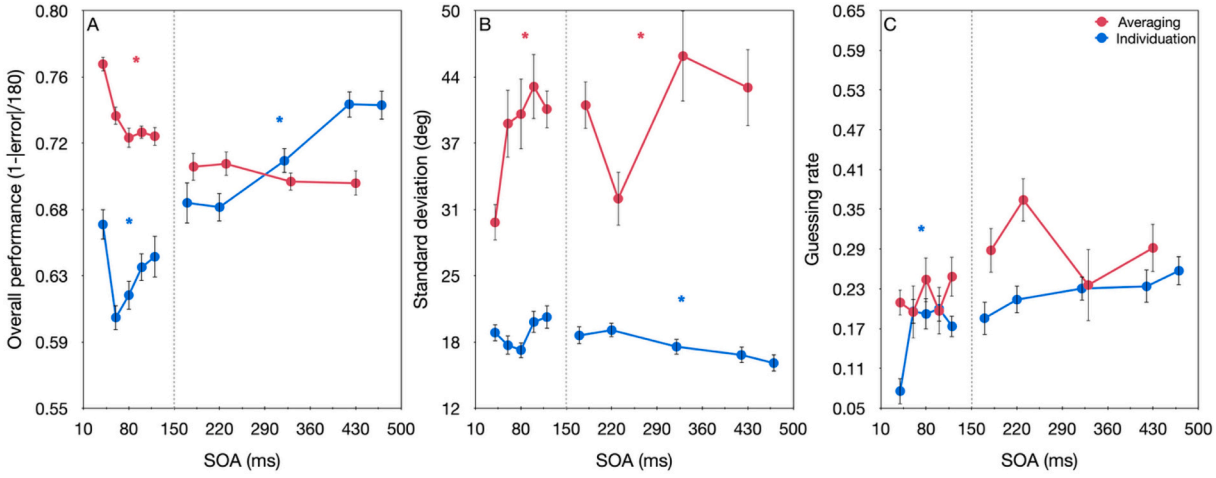

Fig. 2. Estimation error data and the fitted parameters: standard deviation (σ) and guessing rate (γ). (A) Overall performance, (**B**) fitted standard deviation, and **(C)** fitted guessing rate component in the averaging task (red curves) from Experiment 1 (SOA range 40–120 ms) and Experiment 2 (SOA range 180–430 ms). The dotted vertical line marks the border between the SOA ranges of Experiments 1 and 2. For comparison, the blue curves depict the results of the individuation task (Hochmitz et al., 2024; Tkacz-Domb & Yeshurun, 2021) as a function of SOA. Parameter estimates for the individuation data were obtained using the Two-misreport model (see section 2.4.3.1). Error bars represent one standard error. The asterisk indicates a significant SOA effect. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

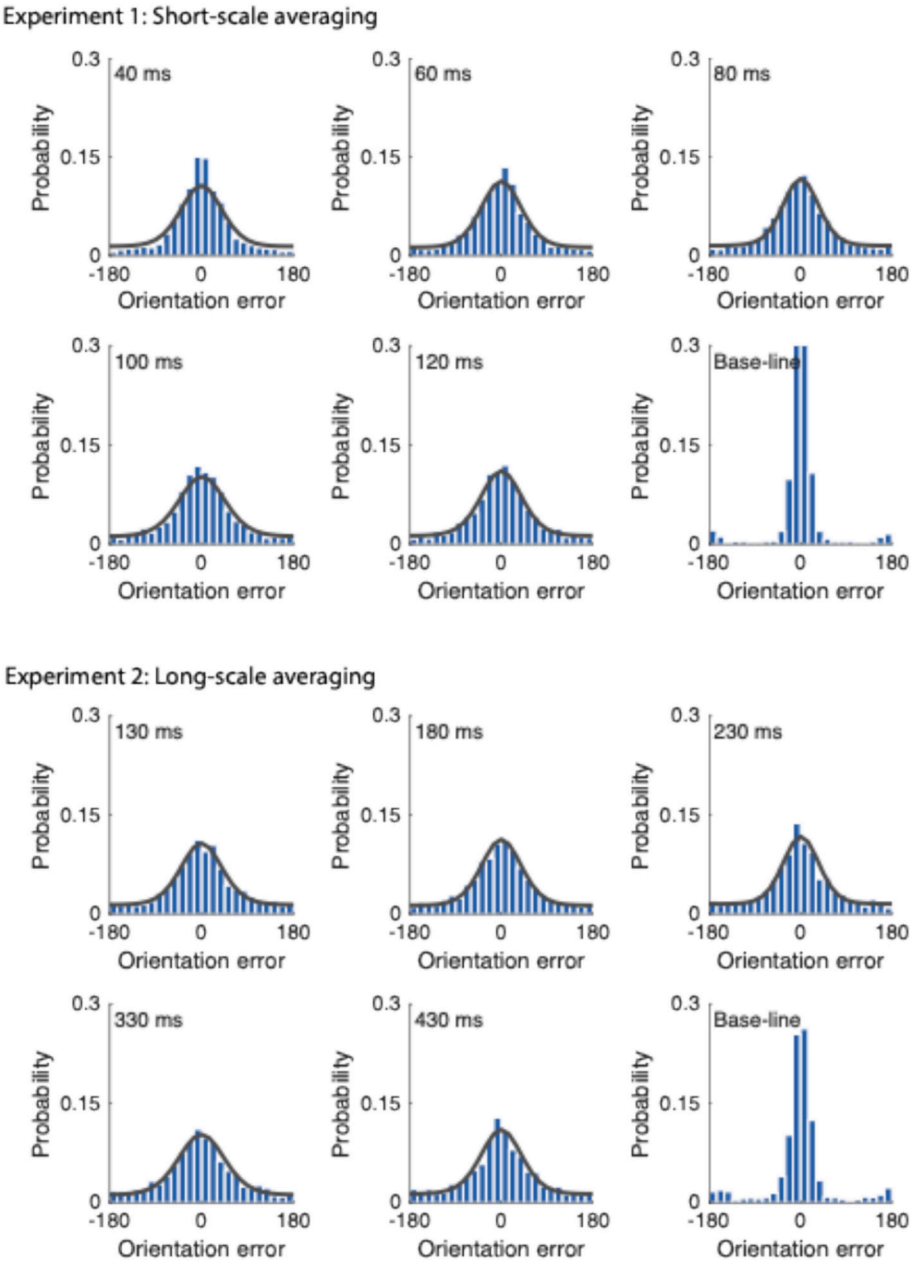

Fig. 4. Mean probability of error distributions for each SOA condition and baseline (a single stimulus) condition of Experiments 1 and 2. Solid lines are the fitting of the Weighted average with guessing model.

difficult to draw a definitive conclusion.

3.2.2.2. Guessing rate (γ). When considering the guessing rate parameter (Fig. 2c, red curve), there was no significant effect of SOA for either of the timescales (Experiment 1: $F(4,72) = 0.645$, $p = .632$, $\eta_p^2 = 0.035$; Experiment 2: $F(3,48) = 1.718$, $p = .176$, $\eta_p^2 = 0.097$).

3.2.2.3. Averaging weights (β_1, β_2, *and* $\beta_3 = 1 - \beta_1 - \beta_2$). Different patterns of effects also emerged for β_1, β_2 and β_3 parameters. For short-scale averaging (Experiment 1), we found a significant effect of SOA for β_2 ($F(4,72) = 2.977$, $p = .025$, $\eta_p^2 = 0.142$) and β_3 ($F(4,72) = 6.702$, $p < .000$, $\eta_p^2 = 0.271$) but not for β_1 ($F(4,72) = 1.807$, $p = .137$, $\eta_p^2 = 0.091$). As can be seen in Fig. 5**: S2, S3** (red curves left panels) the effects for β_2 and β_3 show different patterns: for β_2, a non-monotonic pattern is observed - the weight assigned to the second item first decreases slightly but then increases with SOA. For β_3, however, the assigned weight monotonically decreases with SOA. Additionally, looking at the trend of β_1 (Fig. 5**: S1** red curve left panel) it is evident that overall, the first item is assigned a considerably smaller weight, compared to the second and third items, but its weight increases with SOA.

Long-scale averaging (Experiment 2) exhibited a clearly discernible pattern in the dynamics of weight distribution. Here, there was no significant effect of SOA for any of the three items (β_1: $F(3,48) = 0.787$, $p = .507$, $\eta_p^2 = 0.047$; β_2: $F(3,48) = 0.592$, $p = .624$, $\eta_p^2 = 0.036$; β_3: $F(3,48) = 2.278$, $p = .091$, $\eta_p^2 = 0.125$), with most of the weight assigned to the first and to the second items.

These different patterns of SOA effects in the different temporal scales suggest that averaging involves distinct perceptual interaction at different timescales, likely operating at various processing levels.

3.2.3. Regression analyses of fitted parameters of both temporal scales

The above analysis shows that the pattern of SOA effects observed for

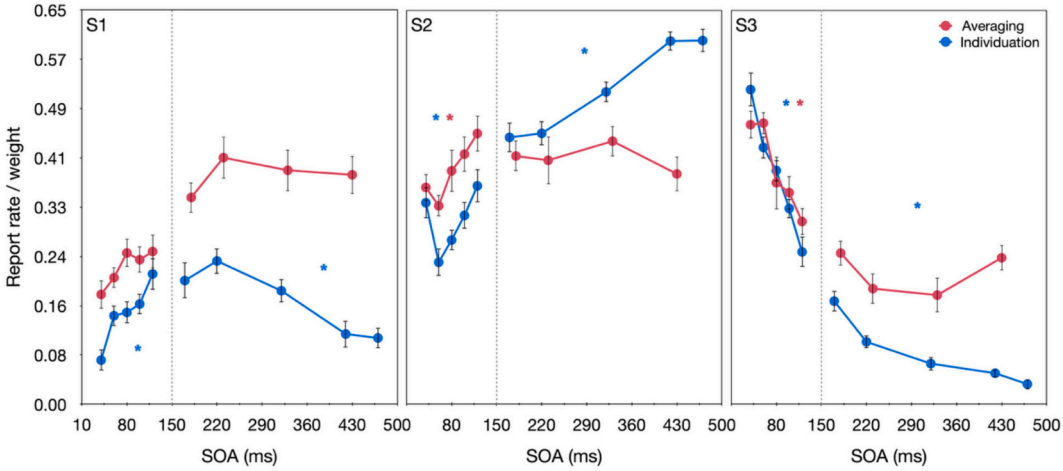

Fig. 5. Fitted weights for each stimulus. Averaging weights (red curves) as a function of SOA for the 1st (S1), 2nd (S2), and 3rd (S3) stimuli in Experiment 1 (SOA range 40–120 ms) and Experiment 2 (SOA range 180–430 ms). The dotted vertical line marks the border between the SOA ranges of Experiments 1 and 2. For comparison, the blue curves represent the report rate of each stimulus in the individuation task (Hochmitz et al., 2024; Tkacz-Domb & Yeshurun, 2021) as a function of SOA. Note that with the individuation task participants had to report the orientation of the stimulus in the middle of the sequence (S2), hence, the larger report rate of S2 represent correct stimulus report, whereas report of S1 and S3 represent misreport (substitution) errors. Parameter estimates for the individuation data were obtained using the Two-misreport model (section 2.4.3.1). Error bars represent one standard error. The asterisk indicates a significant SOA effect. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the model's parameters is very different for short- and long-scale averaging, suggesting that they are mediated by different combinations of perceptual processes (i.e., the relative contribution of perceptual processes changes across temporal scale). In this section, we further test this hypothesis by performing analyses that combine both. Because the two experiments use non-overlapping ranges of SOA, we analyzed the combined data using regression models, for each of the parameters of the Weighted average with guessing model. We first compared which regression model provided the best fit: a model with no breakpoints—assuming that SOA effects over short and long timescales share similar slopes—and piecewise regression models with one or two breakpoints—assuming that the slopes of the SOA effects differ across SOA ranges. If the no-breakpoint model provides the best fit, this would support the conclusion that similar processes mediate averaging across timescales. In contrast, if the models with one or two breakpoints provide the best fit, this would support the conclusion that different mixtures of processes mediate averaging over different timescales. Because this analysis is performed across a continuous SOA range, we also included the 130 ms condition to allow as many data points as possible.

The One-breakpoint model outperformed the other models with all the estimated parameters (see Supplementary Materials, **Supplementary Fig. 5**, for visualization). However, the One-breakpoint regression model for the γ and σ parameters did not reach statistical significance (γ: $F(3, 176) = 2.611$, $p = .053$; σ: $F(3, 176) = 2.538$, $p = .058$) and we,

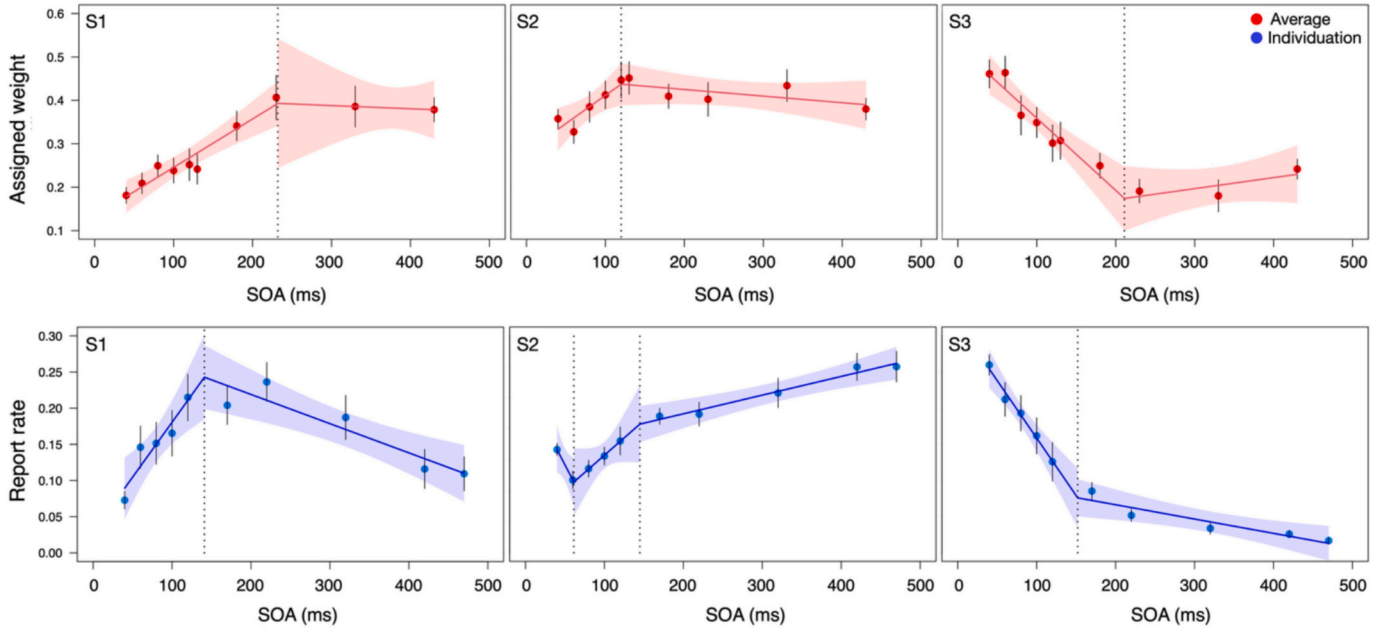

Fig. 6. Upper row: A piecewise regression model fitted to the first (S1), second (S2), and third (S3) items' weights estimated in the averaging task (Experiments 1 & 2). **Lower row:** A piecewise regression model fitted to the estimated report rate of the first (S1), second (S2), and third (S3) items in the individuation task (Hochmitz et al., 2024). In both panels, the vertical dotted line shows the breakpoint (the piecewise regression model that best fitted the S2 report rate of the short-scale individuation data included 2 break points). Error bars represent one standard error. The shaded region corresponds to 95% CIs of the fitted regression.

therefore, continue with the weight parameters β_1, β_2, and β_3 for which the One-breakpoint model was statistically significant (β_1: F(3, 176) = 16.97, $p < .001$; β_2: F(3, 176) = 3.325, $p = .021$; β_3: F(3, 176) = 22.31, p < .001).

3.2.3.1. Averaging weights. The breakpoints obtained by the piecewise regression models are 232 ms, 120 ms, and 211 ms for β_1, β_2, and β_3, respectively (Fig. 6 upper red curves). The regression analyses confirm the results of the separate ANOVAs: The SOA slope with all three parameters was significant for the short-scale (β_1: slope = 0.00116 SE = 0.0002, t = 5.560 $CI_{95\%}$ = 0.0007, 0.0015; β_2: slope = 0.0013 SE = 0.0005, t = 2.573 $CI_{95\%}$ = 0.0003, 0.0023; β_3: slope = −0.0017 SE = 0.0003, t = −5.317 $CI_{95\%}$ = −0.002, −0.001) but was not significant for the long-scale (slope = −0.00007 SE = 0.0005, t = −0.153 $CI_{95\%}$ = −0.0010, 0.0009; slope = −0.0002 SE = 0.0001, t = −1.093 $CI_{95\%}$ = −0.0004, 0.0001; slope = 0.0003 SE = 0.0003, t = 0.970 $CI_{95\%}$ = −0.0003, 0.0008). Critically, with all 3 parameters, the SOA slope of the short-scale was significantly different from the SOA slope of the long-scale (β_1: F(1,176) = 10.517, p = .001; β_2: F(1,176) = 9.200, p = .003; β_3: F(1,176) = 22.458, p < .001). Taken together, these findings support the hypothesis that short-scale and long-scale temporal averaging are mediated by different combinations of perceptual processes that likely operate at different levels. However, the later-than-expected breakpoints that were found for the weights of the first and third items suggest that the dynamics of temporal averaging is different than the dynamics of temporal crowding as with the latter all breakpoints were around 150 ms or earlier (Hochmitz et al., 2024). For comparison, the regression plots of the individuation data are also presented in Fig. 6 (lower blue curves).

3.3. Averaging vs. individuation

Next, we directly compared between the pattern of errors in the averaging task to that in the individuation task. To compare model parameters, we used the fitted parameters of the best model in each task—i.e., the Weighted average with guessing model (Eq. 5) and the Two-misreport model (Eq. 8), fitted to the averaging task and individuation task, respectively. With all parameters, we performed a two-way mixed-design ANOVA, with SOA as a within-subject variable and task (report average vs. middle item) as a between-subject variable.

3.3.1. Short scale

3.3.1.1. Estimation errors. Again, we start with the overall performance scores (Eq. 1; Fig. 2A left panel), as this measure does not depend on any of the mixture models tested here. The ANOVA revealed a significant main effect of task (F(1,33) = 42.036, $p < .0001$, $\eta_p^2 = 0.56$). Participants' overall performance was considerably better in the average task compared with the middle item task. There was also a significant main effect of SOA (F(4,132) = 12.435, $p < .0001$, $\eta_p^2 = 0.27$) and a significant SOA x task interaction (F(4,132) = 3.348, p = .012, η_p^2 = 0.09). As evident in Fig. 2A, a non-monotonic function is observed for the individuation task. Such non-monotonic function is often found with ordinary masking when the target and mask are of comparable strength (reviewed in Breitmeyer & Öğmen, 2006; Francis, 2003). For the average task, however, the overall performance monotonically decreased with the SOA.

Hence, when SOAs are relatively short, reporting the average orientation is easier than reporting the orientation of a specific item. Additionally, SOA modifies overall performance differently when the tasks are different.

3.3.1.2. Model parameters. When looking at the models' parameters further distinctions but also some similarities arise between the two tasks

On the one hand, with the averaging task, the SOA substantially affected the encoding precision (σ), with precision decreasing (σ increasing) as the SOA increased, but it did not affect the guessing rate (γ) (Figs. 2B**&2C**, red curves left panels). In contrast, with the individuation task, an opposite pattern was observed: the SOA mainly affected the guessing rate, which was lowest with the shortest SOA, without affecting the encoding precision (Fig. 2B**&2C**, blue curves left panels). The ANOVAs on the σ and γ parameters corroborate, for the most part, these observations: For the σ parameter, both main effects of task and SOA were significant as well as their interaction (task: F(1,33) = 1.569, p = .219, η_p^2 = 0.05; SOA: F(4,132) = 3.068, p = .019, η_p^2 = 0.09; task x SOA: F(4,132) = 2.662, p = .035, η_p^2 = 0.08). For the γ parameter, neither the main effects nor their interaction were significant (task: F(1,33) = 37.182, $p < .0001$, η_p^2 = 0.53; SOA: F(4,132) = 2.123, p = .081, η_p^2 = 0.06; task x SOA: F(4,132) = 1.962, p = .104, η_p^2 = 0.06).

On the other hand, examining the contribution of each item to the final report, revealed similar patterns across the two tasks (Fig. 5**)**. With both tasks, when the SOAs were very short, the third item dominated the final response. However, as the SOA got longer, the contributions of the first and second items consistently grew at the expense of the third item. The ANOVAs on the *S1–S3* parameters corroborate these observations: For the *S1* and *S2* parameters, both main effects of task and SOA were significant but not their interaction (*S1* – task: F(1,33) = 7.446, p = .010, η_p^2 = 0.18; SOA: F(4,132) = 6.524, p = .0001, η_p^2 = 0.17; task x SOA: F(4,132) = 0.845, p = .499, η_p^2 = 0.03; *S2* – task: F(1,33) = 6.890, p = .013, η_p^2 = 0.17; SOA: F(4,132) = 6.418, p = .0001, η_p^2 = 0.16; task x SOA: F(4,132) = 0.938, p = .444, η_p^2 = 0.03). For the *S3* parameter, only the main effect of SOA was significant (task: F(1,33) = 0.023, p = .881, η_p^2 = 0.00; SOA: F(4,132) = 22.291, $p < .0001$, η_p^2 = 0.40; task x SOA: F(4,132) = 1.600, p = .178, η_p^2 = 0.05).

Thus, the comparison of the two tasks suggests that, under a short temporal scale, averaging and individuation are at least partially mediated by different processes, yet some factors influence performance regardless of the task at hand.

3.3.2. Long scale

With the long temporal scale, the averaging and individuation tasks involved slightly different SOAs (current Experiment 2: 180, 230, 330, 430 ms; Tkacz-Domb & Yeshurun: 170, 220, 320, 420, 470 ms). To compare the two tasks directly within the same ANOVA we included in the analysis only the similar SOAs (180-430 ms, 170-420 ms) and disregarded the minor 10 ms difference between them.

3.3.2.1. Estimation-error data. Starting with overall performance, a distinct pattern of results emerged for averaging and individuation. The ANOVA (SOA x task) revealed a significant main effect of SOA (F(3,96) = 4.097, p = .009, η_p^2 = 0.11) but not of task (F(1,32) = 0.019, p = .892, η_p^2 = 0.00). Importantly, the SOA x task interaction was significant (F(3,96) = 8.711, $p < .0001$, η_p^2 = 0.21) reflecting the fact that reporting the orientation of a single item improved significantly with longer SOAs, whereas reporting the average did not differ with SOA (Fig. 2A right panel).

3.3.2.2. Model parameters. The pattern of SOA effects on the model parameters also differs substantially for the two tasks. With both averaging and individuation, the SOA affected encoding precision, but in a very different manner (Fig. 2B right panel). In the individuation task precision markedly increased (σ decreased) as the SOA increased. In contrast, in the averaging task, there is no clear pattern for this effect suggesting that it bears no true significance. Both tasks show no SOA effect on the guessing rate. When looking at the contribution of each item to the final report (average/middle item) the patterns are again considerably different across the two tasks (Fig. 5: S1-S3 right panels). With the middle item task (blue curves), the report rate of the 2nd item (the target) dominated the final report and further increased with SOA

on the expense of the 1st and 3rd items whose impact decreased with SOA. With the average task (red curves), none of the items' weights differed significantly across the different SOAs. The ANOVAs on the different parameters corroborate these observations: For the σ parameter, the main effect of task was significant as was the task x SOA interaction (task: $F(1,32) = 50.124$, $p < .0001$, $\eta_p^2 = 0.61$; SOA: $F(3,96) = 2.226$, $p = .090$, $\eta_p^2 = 0.07$; task x SOA: $F(4,132) = 3.858$, $p = .012$, $\eta_p^2 = 0.11$). For the γ parameter, neither the main effects nor their interaction were significant (task: $F(1,32) = 2.640$, $p = .114$, $\eta_p^2 = 0.08$; SOA: $F(3,96) = 1.303$, $p = .278$, $\eta_p^2 = 0.04$; task x SOA: $F(3,96) = 1.861$, $p = .141$, $\eta_p^2 = 0.06$). For the *S1* parameter, only the main effect of task was significant (*S1*– task: $F(1,32) = 28.457$, $p < .0001$, $\eta_p^2 = 0.47$; SOA: $F(3,96) = 2.460$, $p = .067$, $\eta_p^2 = 0.07$; task x SOA: $F(3,96) = 1.814$, $p = .150$, $\eta_p^2 = 0.05$). For the *S2* and *S3* parameters, both main effects and their interaction were significant (*S2* – task: $F(1,32) = 5.130$, $p = .030$, $\eta_p^2 = 0.14$; SOA: $F(3,96) = 3.670$, $p = .015$, $\eta_p^2 = 0.10$; task x SOA: $F(3,96) = 6.031$, $p = .001$, $\eta_p^2 = 0.16$; *S3* — task: $F(1,32) = 21.280$, $p = .0001$, $\eta_p^2 = 0.40$; SOA: $F(3,96) = 7.842$, $p = .001$, $\eta_p^2 = 0.20$; task x SOA: $F(3,96) = 3.657$, $p = .015$, $\eta_p^2 = 0.10$).

Thus, in this case, the two tasks show no similarity, suggesting that, under a long temporal scale, averaging and individuation are mostly mediated by different mechanisms.

3.3.3. Comparing breakpoints of piecewise regression

The piecewise regression results also revealed different patterns for averaging and individuation: With individuation, breakpoints were consistently observed around the upper limit of masking (120 ms - 152 ms; Hochmitz et al., 2024), suggesting a relatively stable transition point between masking and temporal crowding effects. Our averaging results, however, revealed a mix of relatively early and late breakpoints across different parameters (120 ms vs ~230 ms; section 3.2.3.1 and Fig. 6). This further suggests that the dynamics of averaging and individuation differ considerably.

4. Discussion

This study aimed to clarify the relationship between averaging and individuation over time, specifically testing whether integration processes can account for temporal interference in the individuation task (i.e., temporal crowding). In addition, this design allow us to test how temporal averaging varies across different temporal scales. To allow a direct comparison between short- and long-scale temporal averaging we used the same orientation estimation task with only the SOA range differentiating the two. Critically, the current study closely followed the stimuli, overall procedure, and analysis framework previously used to measure short- and long-scale individuation. This approach allowed us to both examine the dynamics of temporal averaging and directly contrast averaging and individuation within different scales. A sequence of three orientation stimuli, separated by either short (Experiment 1) or relatively long (Experiment 2) SOAs, was presented to the same peripheral location. The task was to report the average orientation of the sequence of items. Comparisons of the fits of probabilistic mixture models to the error distributions revealed that the Weighted average with guessing model—a model in which averaging weights vary across the items in the sequence—outperformed all other models. This indicates that integration or pooling over time of visual features, such as orientation, varies with the serial position of the features. Moreover, because the averaging weights of the Weighted average with guessing model are comparable to the misreport rate weights of the Two-misreport model (Shechter & Yashar, 2021), we were able to compare the averaging task here with an individuation task using the same stimuli and SOAs (Hochmitz et al., 2024; Tkacz-Domb & Yeshurun, 2021).

4.1. The temporal dynamic of averaging

While studies have examined temporal averaging at various SOAs, they typically used a single fixed SOA, thus one might ask - how do the underlying mechanisms of averaging change across different temporal spans? Our findings demonstrate that temporal averaging can take place across very short and across relatively long SOAs. Critically, further analysis of the fitted parameters revealed that the pattern of averaging weights varies as a function of temporal scale, demonstrating a dissociation in how averaging operates at very fast and longer temporal integration of features. Specifically, for most performance measurements (i.e., overall performance and most model parameters), the SOA played a dominant role in short-scale averaging but had practically no meaningful effect on long-scale averaging. It is possible that with very short inter-item intervals, averaging is mainly mediated by low-level compulsory pooling processes where the actual signals are mixed. With long-scale averaging, such low-level integration processes are less likely to occur, and averaging is no longer affected by the duration of the inter-item interval. This suggests that long-scale averaging is less driven by SOA-dependent low-level interactions and more influenced by higher-level processes, perhaps a more explicit averaging calculation.

The possibility that short- and long-scale averaging are mediated by different combinations of perceptual processes operating at different levels is further supported by the comparison of the weighting profiles across the two timescales. With short-scale averaging, when the SOA is very short, the third item has the highest weight followed by the second item and then the first. This weighting profile could be straightforwardly accounted for by factors such as masking and items' visibility. With the shortest SOAs, the first and second items were poorly registered, because they were backward masked by the following item/s and therefore contributed minimally to the final average report. As the SOA increased their visibility improved, increasing their contribution on the expense of the third item. With long-scale averaging, the long inter-item intervals ensured that all three items were visible, preventing the aforementioned low-level temporal interactions. However, here as well, the weights were not equally distributed across the items. With long-scale averaging, higher weights were assigned to the first and second items than the third item. Hubert-Wallander & Boynton (2015) also found that an item's contribution to the reported average depends on its temporal position in the information stream, creating a temporal weighting profile that varies for different feature dimensions. In their study, the reports of the average object location were more strongly influenced by earlier items, similar to what we found here with stimuli orientation. In contrast, the reports of average size, facial expression, and motion direction were more strongly influenced by later items. Thus, the type of feature that is averaged is important (see also Tong et al., 2019). Notably, when focusing on orientation averaging, our observed primacy effect stands in contrast to the recency effects reported in several prior studies (e.g., Do et al., 2022; Navajas et al., 2017; Wyart et al., 2012). Perhaps these contrasting effects are due to the fact that previous studies employed considerably longer item sequences, allowing more room for recency effects to fully develop. We note that longer sequences might also reveal evidence for rhythmic sampling, as demonstrated by Wyart et al. (2012).

4.2. A dissociation between averaging and individuation

The novel model comparison performed in this study revealed a clear dissociation between averaging and individuation. Regardless of timescale, performance in the averaging task was best explained by the Weighted average with guessing model, suggesting that estimation errors in this case mainly reflect unequal weighting of the items in the average. In contrast, performance in the individuation task was best explained by the Two-misreport model, indicating that estimation errors here reflect imprecise encoding of the target orientation, with occasional misreport of a distractor instead of the target. Both models also account for errors that are merely due to guessing.

Our results also revealed a dissociation between averaging and individuation within each temporal scale. With the short temporal scale, when looking at overall performance, without model fitting, a monotonic decrease with increasing SOAs emerged for the averaging task but a non-monotonic curve was observed for the individuation task. Moreover, it was considerably easier to report the average orientation than the orientation of the middle item. This further implicates the involvement of early low-level integration of signals across time that is beneficial for averaging but not for individuation. Similarly, while the SOA substantially affected precision without affecting the guessing rate for averaging, in the individuation task, it affected the guessing rate without affecting precision. Interestingly, when looking at the contribution of the different items to the final report, the effects of SOA show very similar trends across the two tasks. For both averaging and individuation, the contribution of the first and second items mainly increases with SOA while that of the third item decreases. These similar patterns across the two tasks further suggest that under a short temporal scale performance is affected by low-level factors and processes such as item visibility and temporal integration that occur regardless of the task at hand.

The importance of stimulus visibility as a determinant of performance has long been established. Visibility depends on both the physical characteristics of the input (e.g., luminance/contrast) and its temporal characteristics (e.g., duration, spacing). For brief presentations, Bloch law shows that items' signal strength (energy), and thus their detectability, scales with luminance × duration (time–intensity reciprocity), rather than with duration or luminance alone (Bloch, 1885; Kahneman & Norman, 1964; Di Lollo, 1980). At the same time, visual masking demonstrates that temporally adjacent events can markedly reduce or eliminate a target's visibility (e.g., Breitmeyer, 1984; Gorea, 1987; Breitmeyer & Öğmen, 2000, 2006; Enns & Di Lollo, 2000; Enns & Di Lollo, 2000; Enns, 2004). Masking strength depends strongly on SOA—shorter SOAs increase interference—and when the target and mask have comparable strength, SOA functions can be non-monotonic (e.g., Alpern, 1953; Francis, 2003; Breitmeyer & Öğmen, 2006), as was found in the current study. A recent account, conceptually aligned with the importance of visibility, is the Fidelity-based Integration Model (FIM; Tong & Dubé, 2022a, 2022b). The FIM posits a single sampling mechanism that integrates information from multiple sources in proportion to their relative fidelity (visibility/quality). As representational fidelity varies across items and over trials, each item's influence on the response tracks its current fidelity. Such accounts explain why, at short SOAs, the first and second items contribute less and the third contributes more in both tasks, and why increasing SOA restores the earlier items' contributions across both tasks—e.g., if the first item is hardly registered, then its contribution to the final report should be negligible regardless of whether we are trying to take it into account or to inhibit it. Thus, in the short temporal scale, the two tasks engage different processes, but they also share some common processing mechanisms.

With the long scale, the two tasks are clearly dissociable in every aspect. Whereas with averaging overall performance was not affected by the SOA, with individuation performance increased with SOA. Similarly, unlike the inconsistent pattern found for precision in the averaging task in the current study, long-scale individuation shows a consistent and robust pattern for the encoding precision: precision improves with increasing inter-item intervals (Hochmitz et al., 2024; Sahar & Yeshurun, 2024; Tkacz-Domb & Yeshurun, 2021). Moreover, the pattern of individual items' contribution to the final report across the SOAs considerably varied between the two tasks. In the averaging task, the SOA did not alter the weight distribution, with most weight consistently allocated to the first and second items. Conversely, in the individuation task, the SOA had a notable impact on the report rates of all three items: the report rates for the first and third items decreased with longer inter-item intervals, while the report rate for the second item (the target) increased. These different patterns of effects for the two tasks suggest that performance at this scale is likely mediated by later higher-level processes that differ depending on the task.

Our results also reveal distinct temporal dynamics between averaging and individuation. Specifically, the two processes differ in their transition points across the temporal scales. For individuation, early transitions were previously demonstrated for the contribution of all three items (120 ms - 152 ms; Hochmitz et al., 2024), confirming that the limit of masking is around 100-150 ms (e.g., Breitmeyer & Öğmen, 2000, 2006; Enns & Di Lollo, 2000; Enns, 2004). In contrast, for averaging, we observed later breakpoints for the first and third items (232 ms and 211 ms, respectively). This difference in transition dynamics might be due to the interplay between low-level factors (e.g., item visibility) and high-level task-related factors. In the individuation task, participants must report the middle item while ignoring and perhaps actively inhibiting the first and third items to avoid interference. Such task-induced inhibition would contribute to the early breakpoints observed for the first and third items. Conversely, in the averaging task, participants need to integrate information from all three items. Thus, the averaging task advocates prolongation of items' representation rather than inhibition and this might have led to the pattern of breakpoints we observed.

Finally, the different patterns found for long-scale averaging and individuation also suggest that temporal averaging cannot account for the interference observed with individuating a target from the irrelevant items surrounding it in time. This sets a distinction between temporal crowding—long-lasting interference in the temporal domain—and a simple pooling account that involves compulsory weighted averaging, as some authors have suggested for spatial crowding—interference in the spatial domain (e.g., Greenwood et al., 2009; Parkes et al., 2001). If that was the case with temporal crowding, we would have expected to get a comparable pattern of effects for long-scale averaging (Experiment 2) and temporal crowding (Tkacz-Domb & Yeshurun, 2021). The assertion that temporal crowding is not mediated by averaging is further corroborated by the fact that the Weighted average with guessing model was the best fitting model for the data from the averaging task but not for the temporal crowding data. Importantly, the weighted-average model does not assume that averaging is "compulsory" or computed explicitly; it only determines whether observers' reports are better described as a weighted mean. Indeed, mixture model analysis can capture averaging when observers fail to individuate the target feature from nearby distractors. For example, when compulsory averaging occurs in an individuation task—such as in the case of spatial crowding involving estimation of spatial frequency—the weighted average models outperformed misreport models (Yashar & Carrasco, 2025). Thus, the fact that the weighted average models did not provide better account for the long-scale individuation data than the Two-misreport model allows us to rule out the involvement of compulsory averaging in temporal crowding of orientations. This is, in fact, consistent with spatial crowding, in which orientations are misreported but not averaged (Yashar et al., 2019; Yashar & Carrasco, 2025). Our findings, combined with previous evidence that temporal crowding differs from visual masking (Hochmitz et al., 2024; Sahar & Yeshurun, 2024), portray temporal crowding as a truly unique temporal phenomenon.

4.3. Study limitation

One of the main goals of this study was to compare temporal averaging and individuation using the same stimuli and procedure and this necessitated an experimental design that differs from the typical studies of averaging. Typical studies of averaging across time use considerably larger sequences of items than the one used in the current study, and some have varied systematically the length of the sequence (e.g., Corbett & Oriet, 2011; Hubert-Wallander & Boynton, 2015; Khayat et al., 2023; Khayat & Hochstein, 2018; Tong & Dubé, 2022a, 2022b). Additionally, previous studies usually present items from a narrower range of feature values than the range used here (e.g., Khayat et al., 2023; Khayat & Hochstein, 2018), and several studies have demonstrated that increasing the feature variance reduces averaging performance (e.g., Do et al.,

2022; Navajas et al., 2017; Tong et al., 2019; Wyart et al., 2012). In the current study, the variance was held constant and relatively high, with items randomly picked across 360 orientations. To assess whether this high variability played a critical role in determining our results, we fitted the model to a subset of the data in which the spread of orientations was limited to a range $< 170^{\circ}$ (removing 33 % of trials in each experiment), and plotted the resulting parameter estimates as a function of SOA (Supplementary Materials, **Supplementary Fig. 6**). The observed patterns are highly similar to what we found when all trials were included, confirming the robustness of our main findings.

These differences limit our ability to compare our findings with the existing temporal averaging literature. However, as mentioned above, this approach was necessary for a direct comparison between averaging and individuation over time. We are currently further examining the dynamics of temporal averaging using more typical procedures. We also note that in this study we only investigated explicit temporal averaging within a given trial, and only with orientation estimation and no spatial interference, but implicit temporal averaging, within and across trials, does occur (e.g., Dubé et al., 2014; Khayat & Hochstein, 2018; Khayat et al., 2024; Khayat et al., 2023; Tong & Dubé, 2022a, 2022b), and some studies have found different patterns of results for different features (e.g., Hubert-Wallander & Boynton, 2015; Tong et al., 2019). Thus, future studies should test the role of implicit averaging in the individuation task and whether the pattern of results varies by feature type.

Additionally, when referring to pooling processes, we relied on a descriptive statistic weighted-average model, which captures behavioral performance but does not model underlying temporal integration dynamics or neural population pooling (e.g., Chapman & Denison, 2025; Whitney & Yamanashi Leib, 2018). Thus, although our analysis can assess the contribution of each presented orientation value and their average (pooled) value to behavior, it does not offer a detailed understanding of the underlying processes. Future studies should test such process pooling models to determine how different processes vary in their involvement across short scale and long scale temporal averaging.

Finally, short- and long-scale averaging were measured in separate experiments using different groups of participants, rather than being intermixed within a single session. While this design was used for practical reasons—primarily to avoid excessively long experimental sessions—it introduces some complexity. Specifically, separating the two timescales raises the possibility that participants may have adopted different strategies across experiments, potentially contributing to the observed differences between short and long SOAs. This limitation should be addressed in future work. That said, for the most part, very similar parameters values were observed for the 130 ms and 120 ms conditions, even though they were measured in two different experiments of non-overlapping SOA ranges (Supplementary Materials, **Supplementary Figs. 1, 2**). Thus, if different strategies were adopted in the different experiments, this likely did not play a major role.

5. Conclusions

To summarize, our participants were able to report the average orientation across the range of SOAs examined (40-430 ms), yet their performance was higher with shorter SOAs. Moreover, we found different dynamics for temporal averaging of different timescales. Together, these findings suggest that short- and long-scale averaging differ in how perceptual processes interact. Specifically, short-scale averaging seems to be, at least partially, shaped by compulsory, low-level integration of the actual signals, while long-scale averaging seems to reflect greater influence of higher-level processes, perhaps a more explicit averaging calculation. The comparison of temporal averaging and individuations across timescales also revealed different patterns of results. However, with the short timescale, despite some differences, the contribution of each item in the sequence to the final response was similar for both tasks. This suggests that within a short timescale, low-level factors like item visibility affect performance regardless of the task at hand. In contrast, with the long timescale, no similarities emerged for averaging and individuation, suggesting that when the inter-item temporal intervals are relatively long, performance is dominated by higher-level processes that are tightly related to task demands.

Author note

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

CRediT authorship contribution statement

Ilanit Hochmitz: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Yaffa Yeshurun:** Writing – review & editing, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Amit Yashar:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106374.

Data availability

Data for all experiments have been made publicly available in the Open Science Framework repository:https://osf.io/zkfjq/?view_only=61fbebbc73474b499d11b6aa6253e480.

References

Agaoglu, S., Agaoglu, M. N., Breitmeyer, B., & Ogmen, H. (2015). A statistical perspective to visual masking. *Vision Research, 115*, 23–39.
Alpern, M. (1953). Metacontrast. *Journal of the Optical Society of America, 43*(8), 648–657.
Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392–398.
Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.
Banno, H., & Saiki, J. (2012). Calculation of the mean circle size does not circumvent the bottleneck of crowding. *Journal of Vision, 12*(11), 1–15, 13.
Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10), 1–11, 7.
Bloch, A. M. (1885). Experiences sur la vision. *Comptes Rendus de la Societé de Biologie, 37*(28), 493–495.
Bonneh, Y. S., Sagi, D., & Polat, U. (2007). Spatial and temporal crowding in amblyopia. *Vision Research, 47*(14), 1950–1962.
Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.
Breitmeyer, B. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.
Breitmeyer, B., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision*. USA: Oxford University Press.
Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review, 83*(1), 1–36.
Breitmeyer, B. G., & Öğmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics, 62*, 1572–1595.
Champely, S. (2020). Basic functions for power analysis (R package Version 1.3–0; Computer software). Available at: https://cran.r-project.org/package=pwr.
Chapman, A. F., & Denison, R. N. (2025). A dynamic spatiotemporal normalization model for continuous vision. *bioRxiv*, 2025-03.
Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393 404.
Chung, S. T. (2016). Spatio-temporal properties of letter crowding. *Journal of Vision, 16*(6), 8.
Coates, D. R., Levi, D. M., Touch, P., & Sabesan, R. (2018). Foveal crowding resolved. *Scientific Reports, 8*(1). Article 9177.
Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138*(2), 289–301.
Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109*(1), 75.

Do, J., Eo, K. Y., James, O., Lee, J., & Kim, Y. J. (2022). The representational dynamics of sequential perceptual averaging. *Journal of Neuroscience, 42*(6), 1141–1153.
Dubé, C., Zhou, F., Kahana, M. J., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research, 96*, 8–16.
Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research, 44*(12), 1321–1331.
Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*, 345–352.
Ester, E. F., Klee, D., & Awh, E. (2014a). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 1022–1033.
Ester, E. F., Klee, D., & Awh, E. (2014b). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 1022–1033.
Fisher, N. I. (1995). *Statistical analysis of circular data*. cambridge university press.
Francis, G. (2003). Developing a new quantitative account of backward mask-ing. *Cognitive Psychology, 46*(2), 198–226.
de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences, USA, 108*(32).
Gorea, A. (1987). Masking efficiency as a function of stimulus onset asynchrony for spatial-frequency detection and identification. *Spatial Vision, 2*, 51–60.
Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter like stimuli. *Proceedings of the National Academy of Sciences, USA, 106*(31).
Greenwood, J. A., & Parsons, M. J. (2020). Dissociable effects of visual crowding on the perception of color and motion. *Proceedings of the National Academy of Sciences, 117*(14), 8196–8202.
Hochmitz, I., Abu-Akel, A., & Yeshurun, Y. (2024). Interference across time: Dissociating short from long temporal interference. *Frontiers in Psychology, 15*, 1393065.
Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision, 15*(4), 1–12, 5.
Kahneman, D., & Norman, J. (1964). The time-intensity relation in visual perception as a function of observer's task. *Journal of Experimental Psychology, 68*(3), 215.
Kewan-Khalayly, B., Migó, M., & Yashar, A. (2022). Transient attention equally reduces visual crowding in radial and tangential axes. *Journal of Vision, 22*(9), 1–9, 3.
Khayat, N., Ahissar, M., & Hochstein, S. (2023). Perceptual history biases in serial ensemble representation. *Journal of Vision, 23*(3), 7.
Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision, 18*(9), 23.
Khayat, N., Pavlovskaya, M., & Hochstein, S. (2024). Comparing explicit and implicit ensemble perception: 3 stimulus variables and 3 presentation modes. *Attention, Perception, & Psychophysics, 86*(2), 482–502.
Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in Psychtoolbox-3?*.
Lev, M., & Polat, U. (2015). Space and time in masking and crowding. *Journal of Vision, 15*(13), 1–25, 10.
Lin, Z., Gong, M., & Li, X. (2022). On the relation between crowding and ensemble perception: Examining the role of attention. *PsyCh Journal, 11*(6), 804–813.
Matthews, M. L. (1973). Locus of presentation and the selective masking effect. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 27*(3), 343–349.
Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nature Human Behaviour, 1*(11), 810–818.
Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7).
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.
Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology, 18*(4), 445–451.
Rashal, E., & Yeshurun, Y. (2014). Contrast dissimilarity effects on crowding are not simply another case of target saliency. *Journal of Vision, 14(6):9*, 1–12.
Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11(12):18*, 1–8.
Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science, 2*(1), 437–457.
Sahar, T., & Yeshurun, Y. (2024). Temporal crowding with central vision reveals the fragility of visual representations. *Journal of Experimental Psychology: General, 153*(2), 339–351.
Shechter, A., Medina, S., Share, D. L., & Yashar, A. (2024). Language-universal and script-specific factors in the recognition of letters in visual crowding: The effects of lexicality, hemifield, and transitional probabilities in a right-to-left script. *Cortex, 171*, 319–329.
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*(1), 2116.
Strasburger, H. (2020). Seven myths on crowding and peripheral vision. *I-perception, 11*(3). Article 204166952091305.
Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision, 13*(10), 1–8, 9.
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.
Tkacz-Domb, S., & Yeshurun, Y. (2021). Temporal crowding is a unique phenomenon reflecting impaired target encoding over large temporal intervals. *Psychonomic Bulletin & Review, 28*(6), 1885–1893.
Tong, K., & Dubé, C. (2022a). Modeling mean estimation tasks in within-trial and across-trial contexts. *Attention, Perception, & Psychophysics, 84*(7), 2384–2407.
Tong, K., & Dubé, C. (2022b). A tale of two literatures: A fidelity-based integration account of central tendency bias and serial dependency. *Computational Brain & Behavior, 5*(1), 103–123.
Tong, K., Dubé, C., & Sekuler, R. (2019). What makes a prototype a prototype? Averaging visual features in a sequence. *Attention, Perception, & Psychophysics, 81*, 1962–1978.
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.
Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*(1), 105–129.
Wyart, V., De Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron, 76*(4), 847–858.
Yashar, A., & Carrasco, M. (2025). When periphery rules: Enhanced sampling weights of the visual periphery in crowding across dimensions. *Psychonomic Bulletin & Review, 32*(2), 779–790.
Yashar, A., Wu, X., Chen, J., & Carrasco, M. (2019). Crowding and binding: Not all feature dimensions behave in the same way. *Psychological Science, 30*(10), 1533–1546.
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.
Yeshurun, Y., Rashal, E., & Tkacz-Domb, S. (2015). Temporal crowding and its interplay with spatial crowding. *Journal of Vision, 15(3):11*, 1–16.
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235.