



Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity

Jongsoo Baek¹ · Sang Chul Chong²

Published online: 3 March 2020
© The Psychonomic Society, Inc. 2020

Abstract

The visual system has a limited capacity for dealing with complex and redundant information in a scene. Here, we propose that a distributed attention mode of processing is necessary for coping with this limit, together with a focused attention mode of processing. The distributed attention mode provides a statistical summary of a scene, whereas the focused attention mode provides relevant information for object recognition. In this paper, we claim that a distributed mode of processing is necessary because (1) averaging performance improves with increased set-sizes, (2) even unselected items are likely to contribute to averaging, and (3) the assumption of variable capacity limits in averaging over different set-sizes is not plausible. We then propose how the averaging process can access multiple items over the capacity limit of focused attention. The visual system can represent multiple items as population responses and read out relevant information using the two modes of attention. It can summarize population responses with a broad application of a Gaussian profile (i.e., distributed attention) and represent its peak as the mean. It can focus on relevant population responses with a narrow application of a Gaussian profile (i.e., focused attention) and select important information for object recognition. The two attention modes of processing provide a framework for incorporating two seemingly opposing fields of study (ensemble perception and selective attention) and a unified theory of a coping strategy with our limited capacity.

Keywords Ensemble perception · Focused attention · Distributed attention · Averaging · Population code

Introduction

The visual system has limited capacity (Broadbent, 1958; Luck & Vogel, 1997). One way of utilizing its limited capacity efficiently is through ensemble representation. By summarizing redundant and complex information in a scene, the visual system can rapidly assess the overall properties of the scene and extract its gist (Cavanagh, 2001; Chong & Treisman, 2003), despite the focused attention limits (Cavanagh & Alvarez, 2005; Dux & Marois, 2009; Simons & Levin, 1997). Another way of coping with our limited capacity is focused attention. It selects and processes relevant information

(Carrasco, 2011; Chun, Golomb, & Turk-Browne, 2011), thereby reducing the load on the visual system.

These two separate modes of processing to deal with our capacity limit have been proposed before (Chong & Evans, 2011; Treisman, 2006) because of the following differences. First, the two modes serve different purposes: Ensemble perception is used for extracting the gist of a scene, while focused attention is used for recognizing a few relevant object(s). Second, they deal with the capacity limitation in different ways: Ensemble perception summarizes complex and redundant information, whereas focused attention filters out irrelevant information. Third, they have been suggested to use functionally different pathways: a non-selective pathway for ensemble perception versus a selective pathway for focused attention (Wolfe, Võ, Evans, & Greene, 2011).

Nevertheless, there have been some doubts about whether the ensemble processing mode, separate from focused attention, is necessary. For example, Myczek and Simons (2008) have shown that findings attributed to ensemble processing could be explained by a focused attention mode of sampling a few items. The noise-and-selection model of averaging also

✉ Sang Chul Chong
scchong@yonsei.ac.kr

¹ Yonsei Institute of Convergence Technology, Yonsei University, Seoul, Korea

² Graduate Program in Cognitive Science and Department of Psychology, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul 03722, Korea

showed that observers' averaging performance could well be described by averaging a few selected items within the limited capacity of focused attention (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013). Some studies have found that there are limits to the number of averages computed using the simultaneous-sequential paradigm (Attarha, Moore, & Vecera, 2014) and the pre-cueing method (Huang, 2015). Other studies have found that there are also limits to the number of items included in an average using the ideal observer analysis (Maule & Franklin, 2016) and the set-size manipulation (Ji & Pourtois, 2018).

These findings of capacity-limited ensemble processing led many researchers to investigate how many items contribute to averaging (Allik et al., 2013; Im & Halberda, 2013; Solomon, 2010; Solomon, Morgan, & Chubb, 2011; for a review, see Whitney & Yamanashi Leib, 2018). The estimated number of items varied widely depending on the types of stimuli to be averaged and other variations across studies. It was three in the case of size averaging (Allik et al., 2013; Solomon et al., 2011), but nearly 40 sizes in Lee, Baek, and Chong (2016); more than four for facial expressions (Haberman & Whitney, 2010); and up to 90 orientations (Dakin, 2001). To find a trend in these widely varied estimations, Whitney and Yamanashi Leib (2018) plotted the estimated number of items from 21 studies and found that observers average approximately the square root of the number of items in a display.

Thus, the visual system seems not to use all the available information for averaging. This could be because (1) only selected items contribute to averaging, (2) the averaging process is inaccurate or imprecise, or (3) both. Please note that inaccurate averaging could be due to noise involved with both individual items and the averaging process. Since most studies aimed to determine the capacity limit of ensemble processing, they focused on finding the maximum number of items included in averaging. This led previous studies to conclude that only some items contribute to averaging, whereas others do not contribute to averaging at all (Allik et al., 2013; Myczek & Simons, 2008; Solomon, 2010; Solomon et al., 2011). If we assume that attention selects items that can contribute to averaging (Allik et al., 2013), there is no need to assume two modes of coping with our limited capacity: ensemble perception and focused attention. Only focused attention is an important strategy of coping with our limited capacity.

However, there are reasons to believe that the visual system has an ensemble processing mode, separate from focused attention. First, as we mentioned before, ensemble perception and focused attention serve different purposes, using different methods to cope with limited capacity. Second, some studies have found an improvement in the precision of averaging with

increased set-sizes (Allik et al., 2013¹; Baek & Chong, 2020; Haberman & Whitney, 2010; Lee et al., 2016; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Robitaille & Harris, 2011; but see also Ji & Pourtois, 2018). This is presumably because the noise of individual items could be cancelled out during the averaging task (Baek & Chong, 2020; Galton, 1907; Parkes et al., 2001; Sun & Chong, 2019). On the other hand, performance usually dropped with increased set-sizes in other tasks that required focused attention (e.g., conjunction search: Treisman & Gelade, 1980; visual working memory: Luck & Vogel, 1997), if a set-size exceeded the capacity limit. Indeed, using the same display, Robitaille and Harris (2011) found that averaging performance improved with increased set-sizes, whereas search performance deteriorated. This opposite trend of the set-size effect suggests a separate mode of processing (i.e., ensemble perception).

Third, we think that the visual system is not likely to use only selected items for averaging because even unselected information can contribute to visual processing (Treisman, 1969; Wolford & Morrison, 1980). Consistent with this claim, previous studies have shown that nearly all items contributed to averaging. Chong and his colleagues (Chong, Joo, Emmanouil, & Treisman, 2008) showed that averaging performance dropped when small samples, rather than the entire display, were given. The averaging performance depended on the number of visible items (Joo, Shin, Chong, & Blake, 2009) and on the quality of to-be-averaged items (Jacoby, Kamke, & Mattingley, 2013; Sun & Chong, 2019), and improved with the number of items included (Allik et al., 2013; Baek & Chong, 2020; Haberman & Whitney, 2010; Lee et al., 2016; Parkes et al., 2001; Robitaille & Harris, 2011). Alvarez and Oliva (2008) even found that all presented items had to be included in their average to achieve the observed precision of observers' averaging performance. Finally, to-be-ignored items also contributed to averaging (Oriet & Brand, 2013). These results suggest that the visual system includes far more items for averaging than the limit of focused attention.

Finally, some averaging models (Allik et al., 2013; Dakin et al., 2005; Solomon, 2010; Solomon et al., 2011) and a capacity-estimation method (Rodriguez-Cintron, Wright, Chubb, & Sperling, 2019) assumed a variable capacity when estimating how many items contribute to averaging. In other words, the number of items contributing to averaging can vary across different set-sizes in these studies. We think that observers' intrinsic capacity limit should not vary depending on set-sizes. Baek and Chong (2020) showed that observers' averaging performance can be well described by a model with a fixed attention limit across different set-sizes. In this distributed attention model of averaging, each item contributes to averaging evenly, but its contribution decreases with increasing set-sizes owing to the fixed limits of capacity. This model (Baek & Chong, 2020) outperformed

¹ In this study, thresholds decreased after set-size 2.

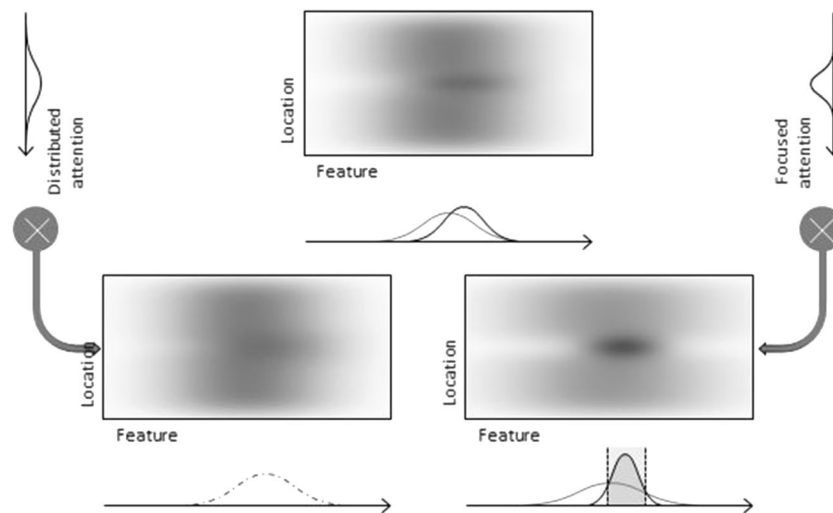


Fig. 1 Two modes of attention (adapted from Cha & Chong, 2018). Each box is an imaginary map of responses to features over locations in one-dimensional space. Darker shades indicate stronger responses. Population responses are depicted below each box in a line-graph. The visual system

can summarize population responses using a distributed attention mode (left) and focus on relevant responses for object recognition using a focused attention mode (right)

the noise-and-selection model with the assumption of variable capacity (Allik et al., 2013) in predicting observers' performance of averaging. Thus, the averaging process is likely to consider all items evenly, rather than only a few selected items.

Thus, it seems that the visual system has two different modes of visual processing to cope with its limited capacity: ensemble perception and selective attention. How then does the visual system access so many items (i.e., over the limit of focused attention) for averaging, given its limited capacity? Hierarchically organized receptive fields in visual processing (Ungerleider & Bell, 2011) and population coding of individual items (Georgopoulos, Schwartz, & Kettner, 1986) may provide an answer to this question. If individual items are represented as a population code in a relevant stage of visual processing, population responses can be easily summarized as a Gaussian-shaped activity profile over them, and we can take the peak value as representing the mean (Fig. 1 bottom left). At the same time, if selection of individual object(s) is necessary for object recognition, the visual system can narrow a Gaussian profile down to relevant responses among population responses (Fig. 1 bottom right).

This idea is schematically demonstrated in Fig. 1. Incoming visual inputs can be represented as population responses that reflect their magnitude and quality depending on locations (Fig. 1 top). If the visual system requires a statistical summary, it can use the distributed attention mode to read it out from population responses (Fig. 1 bottom left). If attention is applied to a broader region in a wide Gaussian profile, responses within a region can be summarized. Likewise, the visual system can use the focused attention mode to select important responses for object recognition (Fig. 1 bottom right). If attention is applied to a specific region in a narrow Gaussian profile, selected responses will increase and thus be

distinguished from others. Previous studies have also suggested the use of a population code to represent statistical summaries (Chong & Treisman, 2003; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2018).

The two different readout mechanisms (distributed and focused attention modes) are different manifestations of a single attention system and can be hierarchically organized. Depending on the purpose of an ongoing task, the visual system flexibly deploys attention in two modes: distributed attention is used for extracting the gist of a scene, while focused attention is used for recognizing specific objects. Consistent with this idea, Cha and Chong (2018) found that observers averaged only relevant orientations for surface perception, suggesting utilization of a different mode of attention to access relevant information. Since attention is involved with multiple stages of visual processing (Kastner & Ungerleider, 2000), these two modes can still interact with each other. For example, an attended object contributes to averaging more than unattended objects (Choi & Chong, 2019; De Fockert & Marchant, 2008).

In conclusion, we propose that the visual system has two separate modes of efficiently managing its limited capacity. Ensemble representation provides a summary of complex and redundant information of a scene, whereas focused attention selects important information from a scene to recognize a few objects. Based on population responses of individual objects, the visual system can read out either a statistical summary for gist perception or crucial information for object recognition.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1A2B5B01070038).

For helpful comments and discussion of this manuscript, we thank Min-Suk Kang and our lab members.

Open Practices Statement There are no data, program code, or experiment associated with the current paper.

References

- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research*, *83*, 25–39. <https://doi.org/10.1016/j.visres.2013.02.018>
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398. <https://doi.org/10.1111/j.1467-9280.2008.02098.x>
- Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1440–1449. <https://doi.org/10.1037/a0036206>
- Baek, J., & Chong, S. C. (2020). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics*, *82*(1), 63–79. <https://doi.org/10.3758/s13414-019-01827-z>
- Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. <https://doi.org/10.1016/j.visres.2011.04.012>
- Cavanagh, P. (2001). Seeing the forest but not the trees. *Nature Neuroscience*, *4*(7), 673–674. <https://doi.org/10.1038/89436>
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*(7), 349–354. <https://doi.org/10.1016/j.tics.2005.05.009>
- Cha, O., & Chong, S. C. (2018). Perceived average orientation reflects effective gist of the surface. *Psychological Science*, *29*(3), 319–327. <https://doi.org/10.1177/0956797617735533>
- Choi, Y. M., & Chong, S. C. (2019). Attending to individual size modulates mean size computation. *Journal of Vision*, *19*(10), 99b. <https://doi.org/10.1167/19.10.99b>
- Chong, S. C., & Evans, K. K. (2011). Distributed versus focused attention (count vs estimate). *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 634–638. <https://doi.org/10.1002/wcs.136>
- Chong, S. C., Joo, S. J., Emmanouil, T. A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, *70*(7), 1327–1334. <https://doi.org/10.3758/PP.70.7.1327>
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. [https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5)
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101. <https://doi.org/10.1146/annurev.psych.093008.100427>
- Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A*, *18*(5), 1016–1026. <https://doi.org/10.1364/JOSAA.18.001016>
- Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research* *45*(24), 3027–3049. <https://doi.org/10.1016/j.visres.2005.07.037>
- De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*(5), 789–794. doi:<https://doi.org/10.3758/PP.70.5.789>
- Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics*, *71*(8), 1683–1700. <https://doi.org/10.3758/APP.71.8.1683>
- Galton, F. (1907). One vote, one value. *Nature*, *75*(1948), 414. <https://doi.org/10.1038/075414a0>
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, *233*(4771), 1416–1419. <https://doi.org/10.1126/science.3749885>
- Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, *72*(7), 1825–1838. <https://doi.org/10.3758/APP.72.7.1825>
- Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2018). Comparing set summary statistics and outlier pop out in vision. *Journal of Vision*, *18*(13), 1–13. <https://doi.org/10.1167/18.13.12>
- Huang, L. (2015). Statistical properties demand as much attention as object features. *PLoS ONE*, *10*(8), e0131191. <https://doi.org/10.1371/journal.pone.0131191>
- Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, *75*(2), 278–286. <https://doi.org/10.3758/s13414-012-0399-4>
- Jacoby, O., Kamke, M. R., & Mattingley, J. B. (2013). Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 233–244. <https://doi.org/10.1037/a0028762>
- Ji, L., & Pouthois, G. (2018). Capacity limitations to extract the mean emotion from multiple facial expressions depend on emotion variance. *Vision Research*, *145*, 39–48. <https://doi.org/10.1016/j.visres.2018.03.007>
- Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision*, *9*(9), 1–12.
- Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341.
- Lee, H., Baek, J., & Chong, S. C. (2016). Perceived magnitude of visual displays: Area, numerosity, and mean size. *Journal of Vision*, *16*(3), 1–11. <https://doi.org/10.1167/16.3.12>
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. <https://doi.org/10.1038/36846>
- Maule, J., & Franklin, A. (2016). Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism. *Journal of the Optical Society of America A*, *33*(3), A22–A29. <https://doi.org/10.1364/JOSAA.33.000A22>
- Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788. <https://doi.org/10.3758/PP.70.5.772>
- Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research*, *79*, 8–16. <https://doi.org/10.1016/j.visres.2012.12.004>
- 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), 739–744. <https://doi.org/10.1038/89532>
- Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, *11*(2), 1–8. <https://doi.org/10.1167/11.12.18>
- Rodriguez-Cintrón, L. M., Wright, C. E., Chubb, C., & Sperling, G. (2019). How can observers use perceived size? Centroid versus mean-size judgments. *Journal of Vision*, *19*(3), 1–14. <https://doi.org/10.1167/19.3.3>
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267. [https://doi.org/10.1016/S1364-6613\(97\)01080-2](https://doi.org/10.1016/S1364-6613(97)01080-2)
- Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*(14), 1–16. <https://doi.org/10.1167/10.14.19>

- Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*(12), 1-11. <https://doi.org/10.1167/11.12.13>
- Sun, J., & Chong, S. C. (2019). Power of averaging: Noise reduction by ensemble coding of multiple faces. *Journal of Experimental Psychology: General*. Advance online publication. <https://doi.org/10.1037/xge0000667>
- Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282-299. <https://doi.org/10.1037/h0027242>
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4-8), 411-443. <https://doi.org/10.1080/13506280500195250>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97-136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Ungerleider, L. G., & Bell, A. H. (2011). Uncovering the visual “alphabet”: Advances in our understanding of object perception. *Vision Research*, *51*(7), 782-799. <https://doi.org/10.1016/j.visres.2010.10.002>
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*, 105-129. <https://doi.org/10.1146/annurev-psych-010416-044232>
- Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77-84. <https://doi.org/10.1016/j.tics.2010.12.001>
- Wolford, G., & Morrison, F. (1980). Processing of unattended visual information. *Memory & Cognition*, *8*(6), 521-527. <https://doi.org/10.3758/BF03213771>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.