The contribution of individual items to the global ensemble representation

Aleksei Iakovlev

Thesis for the degree of Philosophiae Doctor

December 2024

School of Health Sciences

FACULTY OF PSYCHOLOGY

UNIVERSITY OF ICELAND

The contribution of individual items to the global ensemble representation

Aleksei Iakovlev

Thesis for the degree of Philosophiae Doctor

Supervisor(s)

Árni Kristjánsson, Árni Gunnar Ásgeirsson, Igor Utochkin

Doctoral committee

Árni Kristjánsson, Árni Gunnar Ásgeirsson, Igor Utochkin

December 2024

School of Health Sciences

FACULTY OF PSYCHOLOGY

UNIVERSITY OF ICELAND

Hvernig skynjum við hópa ólíkra áreita?

Aleksei Iakovlev

Ritgerð til doktorsgráðu

Leiðbeinandi/leiðbeinendur

Árni Kristjánsson, Árni Gunnar Ásgeirsson, Igor Utochkin

Doktorsnefnd

Árni Kristjánsson, Árni Gunnar Ásgeirsson, Igor Utochkin

Desember 2024

Heilbrigðisvísindasvið
SÁLFRÆÐIDEILD
HÁSKÓLI ÍSLANDS

Thesis for a doctoral degree at the University of Iceland. All right reserved. No part of this publication may be reproduced in any form without the prior permission of the copyright holder.

© Aleksei Iakovlev 2024

ISBN 978-9935-9764-4-4

ORCID: 0000-0002-6256-3752

Reykjavik, Iceland 2024

Abstract

In daily life, people often see groups of similar objects compressed into a single percept. For instance, many people standing together form a crowd, and individual cars moving along the road form traffic. The general impression of traffic speed or crowd mood is perceived without paying much attention to each individual object. Nevertheless, this general impression accurately approximates the whole group. Numerous studies have shown that the visual system can rapidly and accurately represent summary statistics of multiple objects (so-called ensembles) despite observers' poor ability to recall individual items upon which the ensemble representation was built. These findings raise many questions regarding the mechanisms that encode summary statistics, particularly concerning the contribution of individual items to the ensemble representation. Is the ensemble representation built upon the entire group of items or just a subsample? Do sampled items have equal weight in the encoded summary statistics, or do some items determine the ensemble perception while the rest have a minor impact? These questions are addressed in the three papers of this thesis.

In the first paper, we employed computational modeling and behavioral experiments to investigate sampling and robust averaging, which implies larger weight for central elements and lower weight for statistical outliers. Observers adjusted the mean orientation of ensembles with skewed orientation distributions. In line with previous studies, our observers estimated the mean with a systematic bias away from the ensemble mean toward its mode. The bias magnitude increased with the ensemble distribution skewness. This bias can be interpreted in terms of robust averaging: greater skewness means that the extreme elements of the skewed part of the distribution become more deviant from the mean. The more deviant they are, the less their contribution is. As a result, the perceived mean shifts away from the skewed part of the ensemble distribution toward its mode. This pattern was well predicted by the population coding model that utilizes neuron population coding and pooling mechanisms. We conclude, therefore, that this model can be considered a neurally plausible implementation for exhaustive item sampling and robust averaging in ensemble perception.

In the second and third papers, we investigated the effect of saliency on item contribution to ensemble representation (so-called amplification effect) using explicit and implicit reports. Specifically, we tested whether less salient items contribute at all. Our observers performed a mean orientation adjustment task (explicit report) and a visual search for an odd-oriented line among heterogeneous distractors (implicit report). We varied item saliency via size and measured the accuracy and precision of mean orientation estimation in the explicit report and search time in the implicit report. Using explicit reports, we found a strong orientation estimation bias toward the orientation of the more salient items. Still, the observed bias was smaller than what would be produced by subsampling and averaging the most salient items alone. Moreover, less salient items affected the precision of the reported mean. We conclude, therefore, that saliency modulates item contribution to the ensemble representation but does not absolutely determine item sampling. Conversely, we observed no effect of saliency on implicitly reported statistics, suggesting that explicit and implicit ensemble representations may be implemented by different mechanisms.

Keywords:

Ensemble perception; saliency; sampling; robust averaging; feature distribution learning.

Ágrip

Í daglegu lífi sjá fólk oft hópa af svipuðum hlutum samþjappaða í heildarskynjun. Til dæmis mynda mörg einstök áreiti sem mynda hópa, og margir einstakir bílar á ferð eftir veginum mynda umferð. Almenna tilfinningin fyrir umferðarhraða eða stemningu mannfjöldans er skynjuð án þess að mikið sé einblínt á hvern einstakan hlut. Þessi almenna skynjun gefur fremur nákvæma áætlun um einkenni hópsins í heild. Fjöldi rannsókna hefur sýnt að sjónkerfið getur hratt og nákvæmlega áætlað samantektartölur margra hluta (svokölluð áreitasöfn), þrátt fyrir að áhorfendum gangi illa að muna eftir einstökum hlutum sem mynda þessa hópa. Þessar niðurstöður vekja spurningar um þau ferli sem kóða samantektartölur, sérstaklega hvað varðar framlag einstakra hluta til skynjunar á áreitasafninu. Byggir safnið á öllum hlutum hópsins eða aðeins á litlu úrtaki þeirra? Hafa allir hlutir hópsins jafna mikla vigt eða stjórna sumir hlutir skynjun áreitsafnsins á meðan aðrir hafa minni áhrif? Þessum spurningum er svarað í þremur greinum þessarar ritgerðar.

Í fyrstu grein notum við líkön og tilraunir til að sýna að hlutir nálægt miðju safnsins hafa meiri áhrif en tölfræðilegir útlagar. Í samræmi við fyrri rannsóknir áætluðu áhorfendur okkar meðaltalið með kerfisbundnu fráviki frá meðaltali samansafnsins í átt að algengasta gildi þess. Frávikið jókst með aukinni skekkju í dreifingu áreitasafnsins. Meiri skekkja þýðir að öfgahlutir skekkta hluta dreifingarinnar verða meira frábrugðnir meðaltalinu. Því meira frábrugðnir sem þeir eru, því minna leggja þeir til. Afleiðingin er að skynjaða meðaltalið færist frá skekkta hlutanum í dreifingu samansafnsins í átt að tíðasta gildinu. Líkan sem bygggir á kóðun heildarsafna áreita og samvinnu taugafrumna spáði vel fyrir um þetta.

Í annarri og þriðju grein rannsökuðum við áhrif áberandi áreita á skynjun áreitasafna (svokölluð mögnunaráhrif). Við prófuðum hvort minna áberandi hlutir legðu eitthvað til skynjunar áreitasafnanna. Þátttakendur framkvæmdu verkefni þar sem þeir stilltu meðaltalsstefnu og leituðu eftir frábrugðinni stefnu meðal sundurleitra truflunaráreita. Við breyttum áberandi hluta með stærð og mældum nákvæmni stefnumats og leitartíma. Við fundum sterkt frávik í stefnumati í átt að stefnu áberandi hluta þegar þátttakendur svöruðu beint. Enn fremur var frávikið minna en það sem hefði myndast við úrtak og meðaltal áberandi hluta eingöngu. Að auki höfðu minna áberandi hlutir áhrif á nákvæmni skráðrar meðaltalsstefnu. Við ályktum því að áberandi áreiti í áreitasöfnum stýri framlagi hluta til samansafnsskynjunar, en ákveði ekki algerlega þessi áhrif. Aftur á móti fundum við lítil áhrif óbeinnar skynjanar á skynjun áreitasafnanna, sem bendir til þess að skýr og óbein samansafnsskynjun gæti verið útfærð með mismunandi hætti og ólíkum ferlum.

Lykilorð:

Skynjun áreitasafna; áberandi áreit; úrtak; meðaltal; nám á eiginleikum áreitadreifingar.

Acknowledgements

I want to express my deepest gratitude to my supervisor, Dr. Igor Utochkin, for showing me how interesting vision science can be, and for being a good mentor and friend. To my supervisors, Dr. Árni Kristjánsson and Dr. Árni Gunnar Ásgeirsson, for giving me the opportunity to pursue a PhD, for the numerous fruitful meetings, and for their kindness and support. To my friend and co-author, Vladislav Khvostov, for his helpful feedback, productive work, and passion for science that motivates me.

I would like to acknowledge the financial support of the Icelandic Research Fund and the research fund of the University of Iceland for funding my doctoral study and my travel to Iceland. I would like to thank my friend Anton for hosting me in Iceland, for our mental health walks, and for the great songs he introduced me to. I would like to extend my appreciation to all the members of the Icelandic Vision Lab for the supportive and friendly environment they created in the lab and for the awesome events they organized.

I would also like to thank my friends and colleagues for their patience while participating in the long pilot experiments.

Finally, I would like to thank my family for all the love and support they gave me during this journey.

Contents

List of Abbreviations

2AFC - two-alternative forced choice

AIC - Akaike Information Criterion

FDL - Feature Distribution Learning paradigm

IQR - interquartile range

Q1 - first quartile

Q3 - third quartile

RF - receptive field

RT - reaction time

SD – standard deviation

List of Figures

List of Original Papers

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals (I, II, III):

I. Iakovlev, A., & Utochkin, I. (2023). Ensemble averaging: What can we learn from skewed feature distributions?. Journal of Vision, 23(1), 5-5. https://doi.org/10.1167/jov.23.1.5

II. Iakovlev, A., & Utochkin, I. (2021). Roles of saliency and set size in ensemble averaging. Attention, Perception, & Psychophysics, 83, 1251-1262. https://doi.org/10.3758/s13414-020-02089-w

III. Iakovlev, A., Khvostov, V., Ásgeirsson, Á.G., Utochkin, I., Kristjánsson, Á. Amplification from saliency affects explicit but not implicit ensemble representations. Submitted for review in Attention, Perception, & Psychophysics. https://doi.org/10.31234/osf.io/yx4hr

All papers are reprinted by kind permission of the publishers.

Declaration of Contribution

Paper I and II: Aleskei Iakovlev prepared experimental scripts, ran experiments, analyzed data, and contributed to computational modeling and writing the manuscript. Igor Utochkin conceptualized the study, designed experiments, and contributed to computational modeling and writing the manuscript.

Paper III: Aleskei Iakovlev prepared experimental scripts, analyzed data, and contributed to writing the manuscript. Vladislav Khvostov ran experiments and contributed to experiment design, data analysis, writing the manuscript. Árni Kristjánsson, Árni Gunnar Ásgeirsson and Igor Utochkin provided supervision and contributed to experiment design, data analysis, writing the manuscript.

1 Introduction

Let's imagine a citizen who is about to commute to work in order to arrive on time for an important meeting. Before leaving home, they may find it convenient to look out the window and estimate how pedestrians are dressed to decide whether to take warmer or lighter clothing. Certainly, some people outside may be wearing a T-shirt while others may be dressed in a warm coat, but the citizen would still be able to form a general impression. Moreover, the citizen may find it essential to estimate the traffic jams (i.e., the transport speed) to decide whether it would be faster to get to the office by bus or to walk. Unfortunately, the citizen was late, and by the time they arrived, the meeting was already over. The citizen then wondered whether the outcome of the meeting was good or bad. While approaching a large group of colleagues, the citizen would try to estimate whether the colleagues look happy or sad, in other words, whether the meeting went well or not. Fortunately, everyone was happy about the meeting.

The overall traffic, the pedestrians' clothing, and the colleagues' moods were clear to the citizen. Even though such general impressions of groups of objects or people are perceived effortlessly, the visual system must perform complex computations to make this possible. Let's consider the same examples but from a different perspective. The traffic essentially consists of a dozen cars moving at different speeds: while some cars are moving relatively fast, others can be barely moving. Moreover, the overall cars' speed changes over time as streetlights turn green and red. Therefore, to "estimate the traffic," one should average the speed over dozens of moving objects and over some time. Something similar should work for the perception of colleagues' moods. Obviously, each person in the group would express a unique facial expression ranging from deep sadness to extreme happiness. To get a general impression, one should first perceive colleagues' faces over time because sometimes their faces are hidden behind other people, and then average all the sampled facial emotions over time.

This ability to perceive statistical properties of groups of objects is commonly called "ensemble perception" (Alvarez, 2011); the perceived statistics are referred to as "ensemble summary statistics"; and the representation of these statistics is termed "ensemble representation." The earliest experiments on humans' ability to judge statistical properties of multiple objects were conducted in the middle of the 20th century (Peterson & Beach, 1967). However, the one of the founding studies for the contemporary field of ensemble research was conducted by Ariely in 2001. His experiment consisted of two tasks: a member-identification task and a mean-discrimination task. In both tasks, observers were presented with a set of 4 - 16 circles of different sizes followed by a probe circle. In the member-identification task, observers were asked to determine whether the probe circle was a member of the just-presented set. In the mean-discrimination task, observers were asked to evaluate whether the probe circle was larger or smaller compared to the mean circle size of the just-presented set. The results showed that observers were very precise in mean size judgment of a set of up to 16 items, while their ability to identify an individual object in the set was around chance performance. This ability of the visual system to extract the mean of multiple objects has been replicated numerous times in different experiments.

Various studies have shown that the visual system can encode several statistical properties for many different features. In particular, observers can accurately report central tendencies such as the mean (Ariely, 2001; Chong & Treisman, 2003, 2005a), measures of variability such as range or variance (Haberman et al., 2015; Khvostov & Utochkin, 2019; Morgan et al., 2008; Solomon et al., 2011; Suárez-Pinilla et al., 2018), the number of objects, i.e., numerosity (Burr & Ross, 2008; Chong & Evans, 2011; Halberda et al., 2008). Moreover, the ensemble representation includes detailed information about an ensemble feature distribution such as distribution shape, skewness, and bimodality (Chetverikov et al., 2016, 2017a, 2017b). Most of these statistics can be extracted from low- and mid-level features such as orientation (Alvarez & Oliva, 2009; Dakin & Watt, 1997; Parkes et al., 2001), size (Ariely, 2001; Chong & Treisman, 2003), motion speed (Watamaniuk & Duchon, 1992), brightness (Bauer, 2009), temporal frequency of flicker (Kanaya et al., 2018), color (Chetverikov et al., 2017b; Gardelle & Summerfield, 2011; Maule & Franklin, 2015), spatial location, i.e., centroid (Alvarez & Oliva, 2008) and aspect ratio (Elias & Sweeny, 2020). The ability to estimate statistical properties of high-level features has mostly been studied using morphed faces as stimuli. For example, it was shown that observers can estimate mean face emotion (Haberman & Whitney, 2007) and head and gaze direction (Florey et al., 2016; Sweeny & Whitney, 2014). Some studies suggested other high-level features that can be averaged like economic value (Yamanashi Leib et al., 2020) and animacy (Leib et al., 2016), however, such stimuli are rarely used in ensemble perception studies and the ability to average some of them has been doubted (see Khvostov et al., 2020).

Generally, the observer's ability to precisely estimate ensemble summary statistics of a large number of objects while being unable to recall individual items raises many questions. One of the broad questions that is addressed in this dissertation concerns the contribution of individual items to ensemble representation. Do all items contribute to the ensemble representation? Alternatively, is a small subsample selected for further processing? Do processed items weight equally in perceived statistics? Or do some of them have a larger impact on the encoded summary statistics? To address these questions, I will describe in more detail the core concepts related to items' contributions such as subsampling, parallel exhaustive processing, amplification effect, and robust averaging. I will also give an overview of computational models of ensemble perception that account for some of these core concepts. Additionally, I will compare explicit and implicit reports in ensemble studies, since it largely affects the properties of ensemble representation. Then I will summarize the results of three papers that have addressed the question of the contribution of individual items in ensemble representations.

1.1 Explicit and Implicit Representation

Even though the same term “ensemble representation” is used to refer to the representations of different statistics (e.g., mean, variance, feature distribution), different features (e.g., face emotions and orientation), and different types of reports (e.g., explicit and implicit reports), the properties of such representations as well as the mechanisms that implement them could be different. Moreover, even a strong effect observed under one condition cannot be assumed to appear under another. In particular, a growing number of studies suggest fundamental differences between explicitly and implicitly reported statistics. In this section, I will discuss these differences.

An explicit measure involves directly asking observers to report on the ensemble representation, such as the average size. This can be done using methods such as adjustment, where observers adjust a probe circle to match the mean size of a previously presented ensemble, or through two-alternative forced choice (2AFC) tasks, where participants choose which of two circles’ sizes match the mean size of the recently seen set. Additionally, a yes/no response method may be employed, where participants judge whether the probe circle is larger than the mean size of the initially presented set.

Implicit methods assess ensemble representation indirectly, for example, by measuring search times under different conditions, as utilized in the Feature Distribution Learning paradigm (FDL) (for a review, see Chetverikov et al., 2020; Chetverikov & Kristjánsson, 2024). In the FDL, participants search for and locate an odd-one-out target among homogeneous distractors. The trials are structured into blocks where several priming trials precede a test trial. During priming trials, the feature distribution of distractors remains constant, but in the test trial, the features of targets and distractors are swapped by assigning previous distractor features to the target. This swap of target and distractor features results in increased search times due to negative priming (Kristjánsson & Driver, 2008). This increase in search time follows the shape of the distractors’ feature distribution. For example, if the distractor distribution is Gaussian, the closer the distractor that is swapped with the target in the test trial is to the mean of the distractor feature distribution in the priming trials, the longer the search times in the test trial. However, when the distractor distribution is uniform, the search times would be similar unless the test trial target falls within the distractor feature distribution presented during the priming trials (Chetverikov et al., 2017a).

The major difference between these two representations is that the implicit representation has been shown to be richer in detail compared to the explicit one. Hansmann-Roth et al. (2021) tested observers to distinguish ensembles that differed in either mean, variance, or shape of a feature distribution (normal vs. uniform) using explicit and implicit reports. Their study implemented the FDL paradigm with some changes for explicit reports. Trials were arranged in blocks of 3-4 learning trials (i.e., visual search for an odd-one-out target) followed by one test trial. In the implicit report condition, the test trial followed the common FDL procedure described above. In the explicit report condition, two sets of items were presented during the test trials, and the task was to select the set that looked more similar to the distractor set in the preceding search trials. The results suggested that, unlike implicit reports, participants were unable to explicitly distinguish between ensembles with different distribution shapes. The authors concluded that ensembles are encoded in rich detail that can guide behavior implicitly but cannot be accessed for explicit ensemble judgments. Furthermore, the comparison of the precision and uncertainty of explicit and implicit judgments showed no correlation. This led the authors to conclude that implicit and explicit ensemble representations are implemented by partially non-overlapping cognitive mechanisms, mechanisms that may serve different purposes and be used for different tasks.

The study by Hansmann-Roth et al. (2021) utilized different procedures for implicit and explicit reports. Their general conclusion was supported by another study by Khayat, Pavlovskaya, and Hochstein (Khayat et al., 2024) that implemented the same procedure for both explicit and implicit reports. In their study, participants were briefly presented with a set of items that could be either lines with different orientations, circles with different brightness, or circles with different sizes. After that, a test screen with two items was shown. The task was to tell whether the test stimulus was present in a set (i.e., implicit report) or compare the test stimuli with the mean (for instance, by telling whether the test stimulus was larger compared to the ensemble mean, i.e., explicit report). The results indicated that the explicit report was more precise across all judged features compared to the implicit report. The authors reached a conclusion similar to Hansmann-Roth et al. (2021) regarding the different properties and mechanisms of implicit and explicit representations.

Another important difference between implicit and explicit ensemble representation is the ability to encode summary statistics of task-irrelevant ensemble distributions. Studies of explicit ensemble representation have shown that observers can report summary statistics of an ensemble that was either task-irrelevant (Oriet & Brand, 2013; Jackson-Nielsen et al., 2017) or not precued when participants are instructed about the task-relevant ensemble after its presentation (Emmanouil & Treisman, 2008; Halberda et al., 2006; Im & Chong, 2014; Poltoratski & Xu, 2013; Utochkin & Vostrikov, 2017). Conversely, even though the task-relevant feature distribution can be implicitly encoded in rich detail (Chetverikov et al., 2017a), the shape of the task-irrelevant ensemble distribution is unlikely to be implicitly encoded (Hansmann-Roth et al., 2019).

Overall, the comparison between explicitly and implicitly reported statistics reveals fundamental differences between these two types of ensemble representation. These differences indicate a possibility of separate mechanisms for explicit and implicit ensemble representation. Thus, the results of many studies that have utilized explicit reports may not be applicable to implicitly reported statistics. Moreover, the theories and computational models originally developed to explain ensemble averaging may be fruitless in explaining the representation of the whole feature distribution.

1.2 Sampling

To make any contribution to the ensemble representation, an item must be sampled for further processing first. In ensemble studies, observers are usually presented with a large set of stimuli ranging from four to 30 or even more elements (e.g., Dakin et al., 2005). Moreover, the presented stimuli could have very complex features such as facial emotions. The ability of the visual system to extract summary statistics from such stimuli raises questions about the mechanisms that enable such computations. At that time, no known biologically plausible mechanism could quickly and accurately enough average a large set of complex stimuli. Consequently, two theories have been proposed to explain ensemble processing. One theory suggested that all items of an ensemble are sampled and integrated into the ensemble representation, proposing a new mechanism that can account for such computations. Another theory explained ensemble averaging using well-known mechanisms with limited capacity and argued that only a limited subsample of items contributes to the ensemble representation. In this section, I will discuss these two theories in more detail.

Importantly, the debates between these theories mostly relate to explicit ensemble representations. Most computational models and behavioral studies accounting for either parallel exhaustive processing or subsampling were designed to predict explicitly reported statistics, particularly the mean. Studies of implicit ensemble representations have also addressed these debates, suggesting that feature distribution learning requires processing a large number of stimuli, which is unlikely to be performed by a mechanism with limited capacity (Chetverikov et al., 2017a, 2017c). To my knowledge, no studies have yet challenged this conclusion. However, the findings of Chetverikov et al. (2017a, 2017c) do not extend to explicit ensemble judgments.

1.2.1 Parallel Exhaustive Processing

According to the theory of parallel exhaustive processing, all the items of an ensemble contribute to the ensemble representation. All ensemble elements are sampled via distributed attention and then integrated into an ensemble representation with a mechanism that has no restrictions on the number of processed items. However, the more items are processed simultaneously, the less precisely the individual items are encoded (Alvarez, 2011). Initially, it was hypothesized that this mechanism is preattentive (Chong & Treisman, 2003, 2005a), but further studies suggested that some amount of attention is still necessary to consciously perceive ensemble statistics (Jackson-Nielsen et al., 2017), yet broadly distributed across all items (Baek & Chong, 2020b).

The theory of parallel exhaustive processing has mostly been supported by behavioral studies. In particular, these studies have suggested that focused attention and conscious access to the individual elements of an ensemble are not necessarily required to efficiently perform ensemble averaging tasks. For example, it was shown that performance in ensemble averaging tasks improves when distributed attention is required to complete the task compared to focused attention (Chong & Treisman, 2005b; Robitaille & Harris, 2011). Moreover, observers can accurately report summary statistics of an ensemble presented outside the focus of their attention (Alvarez & Oliva, 2008). Additionally, some studies suggested that observers can still accurately report some ensemble statistics, even when conscious access to the ensemble items is disrupted or completely eliminated. For instance, observers can estimate the average despite the crowding effect that disrupts the ability to discriminate individual objects (Solomon, 2010). Moreover, observers suffering from unilateral neglect syndrome (i.e., inability to attend to stimuli presented to the left visual hemifield usually resulting from the lesion of the right parietal cortex) can still encode and report some summary statistics of an ensemble presented in the neglected visual field (Hochstein et al., 2015). Lately, the theory of parallel exhaustive processing was supported in computational modeling approaches (Baek & Chong, 2020a; Utochkin et al., 2024) in addition to the behavioral studies.

1.2.2 Subsampling

Subsampling theory argues that ensemble perception can be explained using the mechanisms of focused attention and working memory. It proposes that only a limited subset of items is selected and processed to approximate the statistics for the entire ensemble. The number of sampled items does not exceed the capacities of focused attention and working memory. This theory was originally proposed by Myczek and Simons (2008). They simulated observers' performance in the ensemble averaging task using a simple computational model of an ideal observer. The aim of this simulation was to determine the minimal number of items needed to be sampled and averaged to match participants' response accuracy. The model had only one free parameter, the number of sampled and averaged items. They used their own data as well as data from previous ensemble studies for computational modeling. Their model suggested that averaging two items is enough to account for participants' response accuracy in ensemble averaging tasks with 2AFC or yes/no responses. However, this model does not include any source of noise, which led to an underestimation of the minimal required number of sampled and averaged items. Further models that included more parameters suggested that a larger number of sampled items was needed. This number could be either constant or depend on the original number of items in an ensemble, for instance, 2-4 objects (Maule & Franklin, 2016), approximately half of the entire ensemble (Allik et al., 2013), or the square root of the presented set of objects (Dakin, 2001; Whitney & Yamanashi Leib, 2018). Moreover, later models have modified the meaning of subsampling in ensemble perception. In the original publication, Myczek and Simons (2008) refer to subsampling in ensemble perception as the selection of a subset of whole items that would be further averaged. Other models, however, describe subsampling as the process of information collection (Allik et al., 2013; Gorea, Belkoura & Solomon, 2014). Furthermore, while Myczek and Simons (2008) proposed their model as a possible mechanism of ensemble perception, some authors explicitly stated that their model serves more as a mathematical algorithm rather than representing the real human observer behavior (Solomon, 2010).

1.2.3 Subsampling vs. Parallel Exhaustive Processing

Both subsampling and parallel exhaustive processing can explain ensemble perception. Moreover, both models have been reconsidered to account for new effects of ensemble perception. For example, parallel exhaustive processing was originally proposed as a distributed attention-based mechanism (Chong & Treisman, 2005a). However, to account for weighting (i.e., when some items contribute more to the ensemble representation), some authors suggested a combination of focused and distributed attention so that distributed attention is responsible for the sampling of all ensemble items while focused attention additionally assigns weights to specific items (Alvarez, 2011; Baek & Chong, 2020a). The subsampling theory initially suggested a fixed number of processed items (Myczek & Simons, 2008). However, some studies showed that with an increase in ensemble set size, observers' speed and accuracy in ensemble averaging tasks improved (Baek & Chong, 2020a; Lee et al., 2016; Robitaille & Harris, 2011). This performance improvement could be more difficult to account for using a model with a fixed number of sampled items; thus, a model with a set-size dependent number of sampled items (e.g., the square root of the presented set of objects) can be used.

Overall, both subsampling and parallel exhaustive processing theories can explain most effects of ensemble perception and can be modified to account for some new effects. Note that behavioral studies are unlikely to prove or disprove either of these theories. Moreover, existing computational models (Baek & Chong, 2020a; Teng et al., 2021; Utochkin et al., 2024) that implement either of these theories can provide equally good predictions of observers' performance in ensemble averaging tasks. Therefore, there is still no consensus regarding whether all items from a set contribute to the explicit ensemble representation or just a limited sample.

1.3 Weighting

Once an item of an ensemble is sampled for further processing, its contribution (or weight) in the ensemble representation can vary depending on different factors. These factors will be discussed in this section. However, before discussing the weighting, it should be noted that it is difficult to distinguish between sampling and weighting using a computational modeling approach. For example, some models have two parameters: the probability of a particular item being sampled and the weight of this item in further averaging. Both parameters range from 0 to 1. The overall contribution of an item across multiple trials could be 0.5 in two cases: when the probability of the item being sampled is 0.5 and the weight in averaging is 1; and when the probability of the item being sampled is 1 and the weight in averaging is 0.5. In other words, a model that implements random subsampling followed by item weighting and their integration with different weights and non-random subsampling (when some items have a larger probability of being sampled) followed by item integration with equal weights could provide the same prediction accuracy of ensemble averaging. Nevertheless, the idea of weighting can be studied separately from the subsampling vs. parallel exhaustive processing discussion. Moreover, well-known models consider weighting as a separate parameter rather than integrating it into the probability of items being sampled (Baek & Chong, 2020a; Gardelle & Summerfield, 2011; Teng et al., 2021).

1.3.1 Weighting Based on Attention Allocation

The effect of attention allocation on ensemble perception has been studied using both top-down and bottom-up attention. Top-down attention was manipulated using secondary tasks (e.g., to remember a particular item of an ensemble); bottom-up attention was manipulated via items' saliency. Following Kanaya et al. (2018), the term 'saliency' is used to refer to how much an item attracts an observer's visual attention. Therefore, 'salient items' are referred to items that are expected to attract visual attention more than others.

Originally, the effect of attention allocation (or attraction) on perceived summary statistics was shown by de Fockert and Marchant (2008). In their study, participants were presented with a set of nine circles of different sizes for 1000 ms. After that, two probe circles appeared. The size of one of the probe circles always matched the mean size of the just presented set, while the size of the other circle was either smaller or larger compared to the mean size. Observers had two tasks: to report the location of either the smallest or the largest circle of a set during the set presentation and to indicate which of the two probe circles had the same size as the mean of the just presented set. The results showed that participants were more likely to choose a larger probe circle if they searched for the largest circle when a set was presented and vice versa. The authors concluded that the reported mean is biased by the features of the attended item. A subsequent study by Choi and Chong (2020) supported this conclusion and specified the effects of attention: first, it increases the apparent size of an item; second, it increases the item's weight in averaging.

The studies described above (Choi & Chong, 2020; de Fockert & Marchant, 2008) utilized top-down deployments of attention (i.e., by the task to search for the smallest or largest item or to report the orientation of the cued item). Following this finding, Kanaya et al. (2018) hypothesized that salient items that naturally attract attention would also modulate the perceived summary statistics. They asked participants to judge the mean size or mean temporal frequency of a set of one, four, eight, or 14 circles. They found that observers systematically overestimated the mean. Moreover, the more items there were in a set, the larger the overestimation bias was. The authors suggested that this growing bias reflects a preferential weighting (or amplification) of a subsample of the most salient items (i.e., items with the largest sizes and highest temporal frequency). The size of the amplified subsample is approximately equal to the square root of the presented number of objects. For instance, when four items are presented, two of the most salient ones are amplified during averaging, resulting in a moderate overestimation bias. When a set of 16 items is presented, four items are amplified, yielding a larger overestimation bias. These findings were also supported using a face emotion averaging task that showed a bias toward strong emotions and its growth as a function of set size (Goldenberg et al., 2021).

1.3.2 Weighting Based on the Deviation from the Central Tendency

Ensemble perception allows the encoding of a group of objects in the form of their different statistical properties, such as the mean. The encoded objects can have large variability. Moreover, some items may significantly deviate from the feature distribution of the rest of the items. Such statistical outliers can largely affect the average if they are weighted equally among the rest of the group. While performing data analysis (for instance, in scientific papers), it is common to exclude statistical outliers to get a more precise estimation of the central tendency. Studies suggest that the visual system can also discount statistical outliers in ensemble averaging tasks.

One of the early studies dedicated to the ability of the visual system to discount statistical outliers was conducted by Haberman and Whitney (2010). In their study, participants performed a facial emotion averaging task. The facial emotions in the set varied from sad to happy. Most of the facial emotions fell within a range of 10 emotional units. However, the set included two outliers which differed by up to 60 emotional units from the mean of a set. The results suggested that the observers' reported average corresponded to the average of faces within the range of 10 emotional units rather than to the global mean, which included two outliers. Nevertheless, the outliers still had a small effect on the perceived mean emotion estimation to some extent. An additional experiment revealed poor observers' awareness of outliers, meaning that outlier discounting wasn't a conscious strategy. Haberman and Whitney (2010) concluded, therefore, that statistical outliers had significantly lower contributions to the perceived mean.

Further behavioral studies and the implementation of computational models suggested a more optimal strategy of ensemble averaging called robust averaging (de Gardelle & Summerfield, 2011; Li et al., 2017). Robust averaging is a weighted averaging strategy that assigns varying weights to items based on their proximity to the mean of a set, giving more weight to items closer to the mean and less weight to those farther away. Some computational models that implement robust averaging, scale the evidence obtained from each sampled element of a set according to its distance from some reference. In ensemble averaging tasks that implement discrete reports (e.g., 2AFC or yes/no responses), the references could be some stimuli (or a feature) that is used to judge the average of the set. For instance, the orientation of a single item that should be compared to the mean orientation of an ensemble (i.e., observers should judge whether the mean orientation of a set was more clockwise or more counterclockwise relative to the orientation of the reference item). In continuous reports (e.g., method of adjustment), other stimuli of an ensemble or ensemble central tendency can be used as a reference in the model (Teng et al., 2021).

Some models of robust averaging in ensemble perception deploy formal mathematical algorithms of weighting rather than biologically plausible mechanisms of the visual system, for instance, an inverse Gaussian loss function (Teng et al., 2021). This modeling approach is good at describing the existing effects but may lack predictive power. Another approach is to implement known mechanisms of the visual system.

A biologically plausible weighting mechanism for ensemble averaging was recently proposed by Utochkin, Choi, and Chong (2024). This model utilizes population coding and synaptic weighting as a neurally plausible mechanism of parallel exhaustive processing and weighting in ensemble averaging. The model includes two levels of the neural network (Fig. 1). The first level is represented by neurons that have small receptive fields (e.g., V1 neurons); the second level consists of neurons that have large receptive fields (e.g., V4 neurons). First, small receptive field neurons encode observed stimuli. Because of the small receptive field size, the population response looks like multiple Gaussians at this level. Then signals spread from the first level to large receptive field neurons of the second level with different synaptic weights that are defined by the tuning curves of these neurons. The output of the second layer neurons is a sum of the signals from the first layer with their synaptic weights. This output is isomorphic to the shape of the physical (stimulus) distribution, meaning that if the feature distribution of observed ensemble is skewed, the large receptive field neuron population response is also skewed.

Overall, ensemble perception raises many questions regarding its mechanisms, particularly concerning item sampling and weighting. Can parallel exhaustive processing and robust averaging be explained using known neuronal mechanisms? Does saliency absolutely determine the probability of an item being sampled, or are less salient items sampled as well? Does saliency affect not only explicit ensemble representation but also implicit representation? These questions are addressed in three papers that will be described in more detail in the following sections of the thesis.

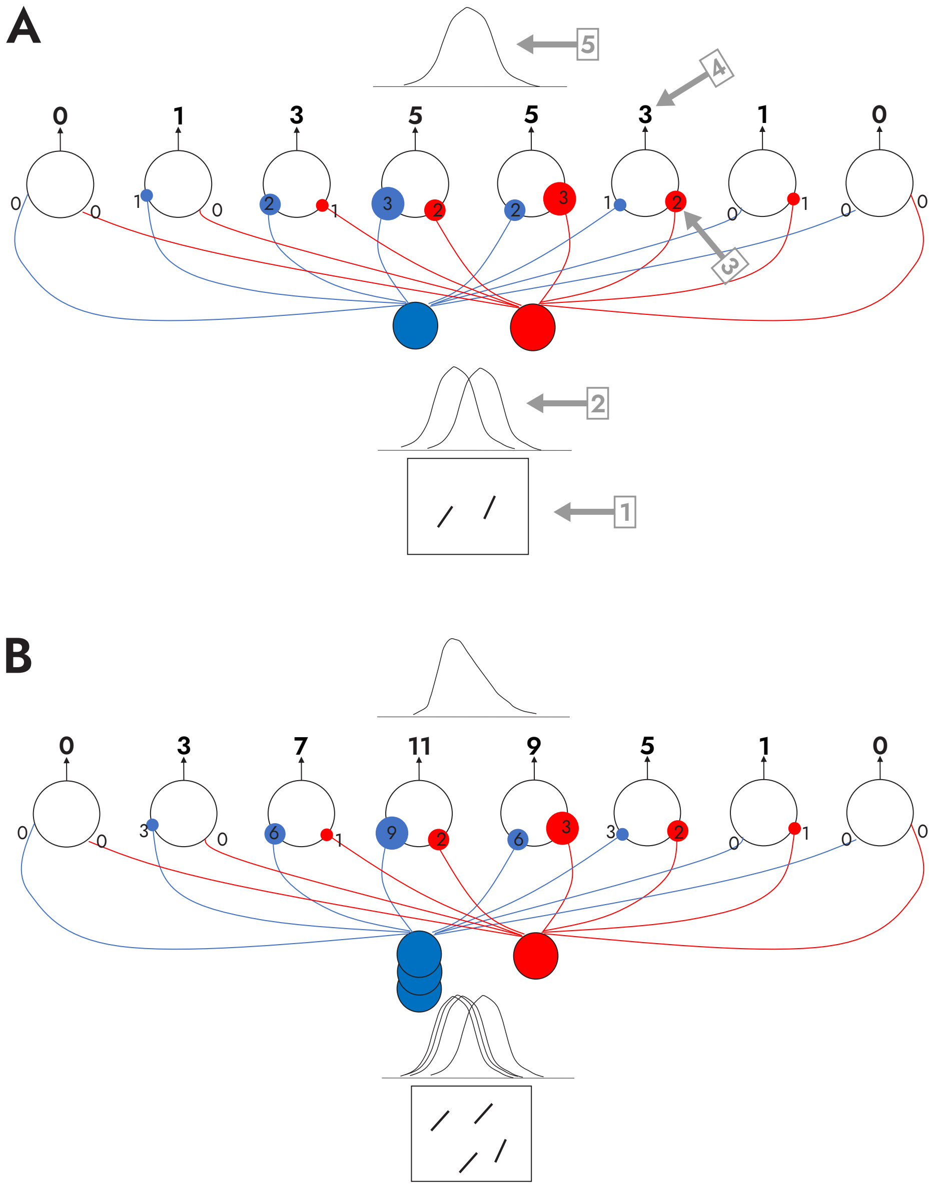


Figure 1. Population coding model of orientation averaging.

(panel A) Presented stimuli (arrow 1) are encoded by the small receptive field which the population response looks like multiple Gaussians (arrow 2). Then activation spreads to large receptive field neurons with different synaptic weights (arrow 3) and summed (arrow 4). The output is the pooled sum of these activations (arrow 5). Peak activation corresponds to the "perceived mean orientation". (panel B) If the physical distribution is skewed, the large receptive field neuron population response is also skewed.

2 Aims

2.1 Paper I

Some recent models of robust averaging implement the subsampling of a limited number of items from the ensemble and then weight them based on their proximity to a central tendency or reference point (de Gardelle & Summerfield, 2011; Li et al., 2017; Teng et al., 2021). Generally, this approach has some limitations. Weighing each sampled element of a set according to its distance from a reference requires encoding a lot of relational information about each element. Furthermore, weighting items relative to the ensemble's central tendency implies that the whole distribution must be available before assigning different weights to individual items. Therefore, implementing subsampling in robust averaging models requires complicated calculations.

In this context, the aim of Paper I was to test a recently proposed model of ensemble perception that utilizes a neurally plausible mechanism of parallel exhaustive processing and robust averaging. Paper I included four behavioral experiments and computational modeling. Experiment 1 examined the impact of distribution skewness on the systematic bias in mean orientation adjustment. Due to constraints imposed by the Covid-19 epidemiological situation on conducting offline experiments, subsequent experiments were conducted online. Experiment 2 replicated Experiment 1 in an online format to validate the results in this new context. Experiment 3 explored whether the systematic bias depended solely on central tendencies (mean and mode) or on the entire shape of the distribution. In Experiment 4, we assessed whether the observed bias was perceptual.

2.2 Paper II

Saliency-based weighting, also known as the amplification effect, was first introduced by Kanaya et al. (2018). In their experiments, participants were asked to perform an ensemble averaging task with features that exhibited salient properties, such as size and temporal frequency. They observed that the perceived mean was biased toward more salient items. The authors concluded that salient items are preferentially weighted compared to other items in a set. The term "preferentially weighted" implies an unspecified fate for less salient items in the original study. This raises several questions: Does only a subsample of the most salient items contribute to the ensemble representation, or are the more salient items amplified alongside others that also process and contribute to the ensemble? Additionally, the study leaves open the question of whether the amplification effect occurs only when saliency is an inherent property of the to-be-averaged dimension, or can a task-irrelevant salient feature of an item also amplify its contribution to the ensemble representation? We addressed these questions in two experiments in Paper II.

2.3 Paper III

In Paper III, we continued to investigate the impact of saliency on ensemble representation. To our knowledge, the amplification effect has previously been demonstrated only in ensemble averaging tasks, such as in size averaging tasks. We explored whether saliency influences not only the perceived mean but also other summary statistics, such as the entire shape of the feature distribution. It has been suggested that the representation of the entire feature distribution can only be accessed using implicit reports (Hansmann-Roth et al., 2021). Consequently, in two experiments we addressed two interconnected questions: Does saliency affect the entire shape of the feature distribution, and can the amplification effect be revealed in tasks involving implicit measures? Experiment 1 was aimed at replicating the amplification effect in the orientation ensemble averaging task using the method of adjustment (i.e. explicit report). Experiment 2 was designed to examine the amplification effect for the same stimuli but using Feature Distribution Learning paradigm (i.e. implicit report).

3 Materials and Methods

Stimuli in all studies were created using PsychoPy 3 (Peirce et al., 2019) and presented against a grey background. In Paper I (Experiment 1) and Paper II, stimuli were displayed on a standard VGA monitor with a refresh rate of 75 Hz and a spatial resolution of 1600 × 1200 pixels. Stimuli in Paper III were presented on a standard LCD monitor with the refresh rate of 60 Hz and resolution of 1920×1080 pixels. In Paper I (Experiments 2-4), stimuli were presented using the online platform Pavlovia (https://pavlovia.org) and participants completed the experiments on their personal computers. All participants reported having normal or corrected-to-normal vision and no neurological problems. Before the experiment, all participants signed an informed consent in paper or electronic form.

3.1 Paper I

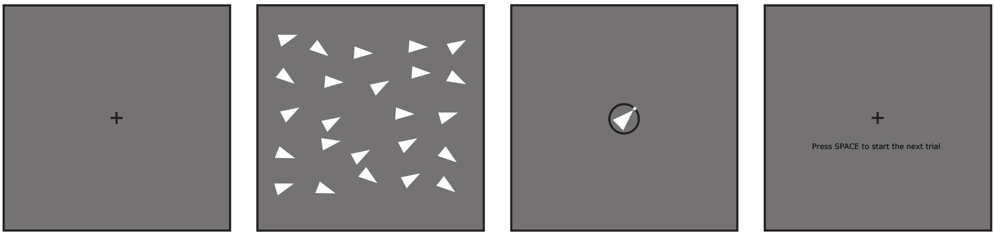
Twenty experienced observers (students of HSE University, five females; average age = 26.5, SD = 6.8) participated in Experiment 1. Participants for Experiments 2, 3, and 4 were recruited through the online platform Prolific (www.prolific.ac). Fifty-one observers (22 female, average age = 25.7 years, SD = 9.5) participated in Experiment 2; fifty observers (28 female, average age = 24.2 years, SD = 5.5) participated in Experiment 3; fifty observers (33 females, average age = 25.7 years, SD = 9.5) participated in Experiment 4. None of the observers participated in more than one experiment. Due to low response accuracy, data were excluded for: two participants in Experiment 2, four participants in Experiment 3, and five participants in Experiment 4.

3.1.1 Paper I: Stimuli and Procedure

At the beginning of each trial, a fixation cross was displayed for 500 ms, followed by a stimulus set presented for an additional 500 ms. The stimulus set consisted of 24 white triangles with varying orientations in Experiments 1 and 2, and 36 white triangles in Experiments 3 and 4. The orientation distribution was skewed, defined as the difference between the mode and the arithmetic mean of the distribution. The skewness magnitude and The skewness magnitude and sign were gradually varied. were gradually varied. For Experiments 1 and 2, the orientations of the triangles in the base distributions ranged from 0° to 63°, with increments of 9°. For Experiments 3 and 4, they ranged from 0° to 112°, with increments of 16°. The orientation distribution could be rotated by adding a random integer from 1° to 360° to each orientation in every trial, ensuring random assignment of the mean orientation. Depending on the skewness of the distribution, the ratio of triangles with varying orientations was adjusted. We used skewed distributions in Paper I because ensemble averaging of skewed distributions has been shown to be sensitive to how individual items are sampled and weighted (Kim & Chong, 2020; Teng et al., 2021).

Subsequently, a probe stimulus was displayed and remained visible until the participant responded. In Experiments 1 to 3, a single triangle was used as the probe stimulus. In Experiment 4, the probe stimulus comprised 36 triangles, whose skewness either matched or was the opposite of the skewness of the previously viewed set. Participants were instructed to adjust the mean orientation of the probe triangle, or the mean orientation of the probe set, to match the mean orientation of the previously viewed set. To make adjustments, participants were required to press and hold the mouse button while moving the mouse cursor around the center of the screen. The orientation of the probe triangle or the mean orientation of the probe set followed the movement of the mouse cursor. Participants confirmed their responses by pressing the spacebar. Before the experimental block, participants completed a training session during which they received feedback on the accuracy of their responses. No feedback was provided during the experimental block.

In Experiment 4, we implemented a probe stimulus comprised of a set instead of a single triangle. This was to test whether the bias is associated with the representation of the whole ensemble distribution or just reflects a response bias. The mean estimation bias elicited by the encoded feature distribution should be diminished when observers adjusted the mean orientation with a probe stimulus set having the same distribution skewness as the test stimuli. Conversely, the bias should increase when the test and probe stimuli sets had opposite skewness.



Fixation cross (500 ms)

Sample set (500 ms)

Probe stimulus (until response)

Intertrial interval (until response)

Figure 2. Stimuli and Procedure of Paper I (Experiments 1 and 2).

3.1.2 Paper I: Analysis

Participants' data analysis involved calculating the adjustment error as the difference between a participant's response and the correct circular mean orientation (Error = Response – Correct Mean). The error could range from –180° to 180°. We then calculated the circular mean of the error distribution for each condition and observer. This mean error was used as a measure of bias. A positive bias indicates that participants systematically adjusted the mean orientation more clockwise relative to the actual mean orientation. A negative bias shows that participants tended to give a more counterclockwise response relative to the correct answer. We used a linear mixed-effects model to analyze changes in bias as a function of the skewness of the feature distribution. This model incorporated the fixed effect of skewness and the random effects of participant identity, slope, and intercept, using the lme4 package for R (Bates, Mächler, Bolker, & Walker, 2014). P-values for the regression were calculated using Satterthwaite's (1946) method.

In Experiments 1-3, we excluded the entire participant's data if the SD of their error distributions exceeded 60°. Since Experiment 4 was more difficult for the participants, we increased the criteria for data exclusion from 60° to 80°. We also excluded individual trials that exceeded three SDs of the entire error distribution for a given observer across all conditions in all four experiments. Using these criteria, in Experiment 1, we excluded 0.7% of trials, with no participants removed. In Experiment 2, we excluded 1% of trials and the data of two participants, resulting in the analysis of data from 49 participants. In Experiment 3, we excluded 1.3% of trials and the data of four participants, resulting in the analysis of data from 46 participants. In Experiment 4, we excluded 1.1% of trials and the data of five participants, resulting in the analysis of data from 45 participants.

Modeling predictions of the magnitude and direction of systematic error were made through Monte Carlo simulations based on the population coding model (as in Utochkin et al., 2024, with some important modifications) and the loss function model (Teng et al., 2021). The free parameters of the first model were the width of the tuning function for summing neurons (σ tuning) and the constant error (a general tendency to choose an answer with a clockwise or counterclockwise bias). The free parameters of the second model included the standard deviation of the inverted Gaussian loss function (σ loss), the number of randomly selected ensemble elements sufficient to estimate the mean with a given accuracy (N), and the constant error (for more detailed model description, see Teng et al., 2021). The best parameters for each model were chosen by maximum likelihood estimation. Model comparison was performed using the Akaike Information Criterion (AIC).

3.2 Paper II

Twenty-six students from HSE University (23 females, average age = 19.9, SD = 1.3) participated in Experiment 1, and twenty-six students from HSE University (21 females, average age = 19.8, SD = 1.3) took part in Experiment 2. Ten of these observers participated in both experiments. Due to low performance, the data of one participant were excluded from Experiment 2.

3.2.1 Paper II: Stimuli and Procedure

Each trial started with the presentation of a fixation cross for 500 ms, followed by a set of triangles displayed for an additional 500 ms. The stimulus sets comprised four, eight, or 16 white isosceles triangles that varied in size and orientation, maintaining a 2:1 height-to-width ratio. The orientations of the triangles were set at -30°, -10°, +10°, or +30° relative to a randomly selected mean orientation, which ranged from 1° to 360° for each trial and was never included in the set. The triangles were either uniform in size or divided into two groups: small and large. The sizes were determined by randomly selecting a mean height between 0.63° and 1.26° of visual angle, with large items at 1.3 times the mean height and small items at 0.7 times the mean height. In conditions with triangles of the same size, all individual triangles matched the mean height while preserving the 2:1 height-to-width ratio.

Immediately following this, a randomly oriented probe triangle, surrounded by a black adjustment ring with a white slider, appeared at the center of the screen. This display remained until the participant confirmed their response. The slider's position on the ring indicated the direction of the probe triangle's apex. Participants were instructed to adjust the probe triangle's orientation to align with the mean orientation of the just-presented set. Adjustments to the probe triangle's orientation could be made by dragging the slider or clicking with the mouse. After completing the adjustment, participants pressed the spacebar to confirm their response. Before the next trial, a blank screen appeared, instructing participants to press the spacebar when ready to continue, thereby allowing for a comfortable pace and breaks as needed. No instructions regarding the size of the triangles were given. Prior to the experimental session, participants campleated a training session to familiarize themselves with the stimuli and procedures. During the training session, feedback on the accuracy of their responses was provided after each trial.

3.2.2 Paper II: Analysis

We calculated the adjustment error using a formula similar to that in Paper I (Error = Response – Correct), which could range between -180° and 180°. The resulting error distributions for each participant and each experimental condition were used to estimate two parameters: the mean, as a measure of systematic bias, and the standard deviation (SD), as a measure of representational imprecision. A positive bias suggests that participants tended to adjust the mean orientation in a more clockwise direction compared to the true mean orientation. Conversely, a negative bias implies that participants' responses were generally more counterclockwise than the actual correct orientation. We compared the bias and SD between the conditions using repeated measures ANOVA and post-hoc t-tests.

3.3 Paper III

Twenty participants (four females; average age = 29.45 years, SD = 9.65) took part in the experiment.

3.3.1 Paper III: Stimuli and Procedure

In this experiment, participants completed two tasks: a mean orientation adjustment task and a visual search task for an odd-one target among heterogeneous distractors. Each trial in both tasks began with a 500 ms presentation of a fixation cross, followed by a set of lines. In the adjustment task, the lines were displayed for an additional 500 ms, while in the visual search task, they remained visible until the participant responded. The stimulus sets comprised 36 white lines varying in size and orientation. The orientations were uniformly distributed, with individual line orientations set at -30°, -18°, -6°, 6°, 18°, and 30° from the mean. The mean orientation was randomly chosen between 1° and 180° for each trial. Lines were either 40×5 or 64×8 pixels in size The size and orientation could be correlated; for instance, larger items could have a more clockwise orientation relative to the mean orientation of the ensemble (i.e., lines with orientations of 6°, 18°, and 30° from the mean), or vice versa, larger items could have a more counterclockwise orientation (i.e., -30°, -18°, -6° from the mean). Alternatively, size and orientation could be uncorrelated, meaning that both larger and smaller lines could have any orientation. The key difference between the tasks was that in the visual search task, one line was oriented 90° from the mean orientation, serving as the target, with the remaining thirty-five lines acting as distractors.

In the visual search task, participants were instructed to locate the oddly-oriented line and determine whether it was in the upper or lower part of the screen by pressing the corresponding arrow button. Feedback was provided after each response. Trials were separated by a 100 ms interstimulus interval and proceeded automatically. However, every 60 trials, participants were given the option to take a break or continue by pressing a button.

In the adjustment task, following a 500 ms stimulus presentation, a probe stimulus appeared and remained onscreen until the participant confirmed their response. The probe consisted of a randomly oriented line at the screen's center, encircled by a black adjustment ring with a white slider. The slider's position indicated the direction of the probe line's apex. Participants adjusted the probe line's orientation to match the mean orientation of the previously viewed set by dragging the slider. The response was confirmed by pressing the spacebar. Additional confirmation was required to start the next trial, allowing breaks between trials as needed. Feedback on the accuracy of participants' responses was provided only during a brief training session.

3.3.2 Paper III: Analysis

In the adjustment task, we analyzed the orientation estimation bias, drawing on methods similar to those used in Paper I and Paper II, with the modification that the bias could range between -90° and 90°. In the visual search task, we assessed the average reaction time (RT). For both tasks, we used repeated measures ANOVA and post-hoc t-tests for our analyses. We excluded individual trials if the response exceeded the third quartile (Q3) plus 1.5 times the interquartile range (IQR) or fell below the first quartile (Q1) minus 1.5 times the IQR. Additionally, we excluded trials from the adjustment task with signed errors outside the [+60°, -60°] range and from the visual search task with RTs shorter than 200 ms.

4 Results

The detailed results and statistical analyses are elaborated in the papers themselves, provided in the appendix. This section will provide a summary of the most important findings from each study.

4.1 Paper I

Behavioral Data: Observers' responses were systematically biased away from the mean and toward the mode of the skewed distribution. The greater the distance between the mean and mode, the stronger the bias. This effect was consistent across four experiments, which varied in testing conditions (offline in-lab vs. online), shapes of feature distributions, and types of probe stimuli (single object or a set of objects with different skewness). Results from Experiment 3 indicated that distributions with the same skewness but different shapes led to biases of varying magnitudes. In Experiment 4, the systematic bias diminished when both the test and probe stimuli had the same distribution shape (e.g., both skewed to the right). However, when the test and probe stimulus distributions had opposing skewness directions, the systematic bias increased.

Modeling: We proposed a biologically plausible model to explain these patterns, based on fundamental and well-established mechanisms of visual system organization, such as population coding and spatial pooling of local neural responses by neurons with large receptive fields. Our model demonstrated high predictive accuracy in Monte Carlo computational simulations. Furthermore, we compared our model with another recent model by Teng et al. (2021), which employs a different mechanism for the ensemble averaging. Despite the fundamental differences in their underlying principles, both models accurately predicted systematic biases in the averaging of skewed distributions.

4.2 Paper II

Experiment 1 demonstrated that the perceived mean orientation was significantly biased towards the orientation of more salient items. Specifically, participants showed a clockwise bias in their adjustments of the mean orientation when the more salient items were oriented more clockwise relative to the set's overall mean orientation; a counterclockwise bias was observed when the more salient items were oriented more counterclockwise. Moreover, this bias increased as a function of the number of presented items. Experiment 2 demonstrated that the precision of averaging improved when the orientations of larger items were closer to the overall set's mean orientation and deteriorated when the orientation of these more salient items was far from the mean. Additionally, the presence or absence of less salient items influenced the precision of the perceived mean.

4.3 Paper III

Experiment 1 demonstrated a strong amplification effect for an explicit report in the ensemble averaging task. This effect was similar to that observed in Experiment 1 of Paper II. However, the results of Experiment 2 showed that saliency had no effect on the implicitly reported summary statistics.

5 Discussion

The studies presented in this thesis were dedicated to understanding of sampling and weighting in ensemble encoding. First, we examined a computational model that suggests a neurally plausible mechanism for parallel exhaustive item processing and robust averaging in ensemble perception (Paper I). Second, we investigated the so-called amplification effect from salient stimuli upon ensemble perception using explicit reports (Paper II). Finally, we compared implicit and explicit reports to reveal to what degree they are affected by amplification (Paper III), elaborating on the differences between explicitly and implicitly reported summary statistics.

5.1 Robust Averaging

In Paper I, we tested the predictions of the recently proposed population coding model (Utochkin et al., 2024) using a mean estimation task for ensembles with skewed feature distributions. We observed a systematic mean estimation bias away from the ensemble mean and toward its mode. The bias magnitude increased with the distance between the ensemble orientation distribution's mean and mode. These findings are consistent with previous studies of ensemble averaging with skewed feature distributions (Teng et al., 2021; Webb et al., 2007, 2010). Additionally, we found that the bias magnitude was affected by changes in the feature distribution shape, even when the physical mean and mode remained the same. This distribution-dependent change in bias supports the hypothesis that the perceived central tendency is calculated from the neuron population response, built upon the physical ensemble distribution, rather than determined by summary statistics of central tendency alone (Baek & Chong, 2020b; Brezis et al., 2017; Dakin et al., 2005; Haberman & Whitney, 2012; Utochkin et al., 2024; Webb et al., 2007, 2010). Even though the ensemble feature distribution cannot be judged explicitly (Hansmann-Roth et al., 2021), its importance for the perceived average was shown in several studies (Kim & Chong, 2020; Michael et al., 2014). Moreover, we found that the mean estimation bias diminished when observers adjusted the mean orientation with a probe stimulus consisting of an ensemble having the same distribution skewness as the test stimuli. Conversely, the adjustment bias increased when the test and probe stimuli sets had opposite skewness. These findings suggest the observed effect is a result of perceptual processes rather than a response bias.

In addition to the behavioral study, we tested a recently proposed population coding model of ensemble perception (Utochkin et al., 2024) and compared the predictions of this model with the loss function model suggested by Teng et al. (2021). Both models implement robust averaging, meaning that statistical outliers contribute less to mean estimation. In terms of robust averaging, the observed biases caused by asymmetrical feature distributions result from the lower contribution of outlying items (i.e., items located in the longer tail of the distribution), which shifts the perceived mean away from the longer "tail" of the distribution. Although both models employ a similar concept (i.e., robust averaging), a direct comparison would be incorrect since these models correspond to two different levels of Marr's (1982) model classification. Teng et al. (2021) focus on a formal mathematical algorithm of ensemble averaging. Although the loss function model could be potentially explained by some neuronal mechanisms (e.g., feedback processes, Epstein et al., 2020), Teng et al. (2021) do not consider them. In contrast, the population coding model focuses on the implementation of neurally plausible mechanisms with their constraints. Therefore, we compared the models to evaluate whether the population coding model provides a good fit to the data rather than to reveal which model is superior. We found that both models accurately predicted observers' mean orientation estimation bias, suggesting that the population coding model can be considered a neural implementation of parallel exhaustive processing and robust averaging in ensemble perception.

Overall, different experiments on ensemble perception with skewed feature distributions have demonstrated a systematic bias away from the ensemble mean toward its mode. Potentially, such biased perception of the mean in skewed distributions could be superior to a precise one. From a statistical perspective, it is more reliable to use the median (which lies in between the mean and the mode) as a measure of central tendency in skewed distributions since the mean can be misleading. For instance, salaries in some countries exhibit a right-skewed distribution, meaning that the "mean" salary is much higher than most people's incomes. Meanwhile, the median salary provides a more reliable estimation of the citizens' wealth. Similarly, in perceptual tasks, the biased mean estimation of ensembles with skewed distribution could be more reliable than unbiased.

5.2 Amplification Effect

We examined the effect of saliency amplification in Paper II using explicit reports and in Paper III using explicit and implicit reports. In both papers, we manipulated saliency through item size in a set of oriented lines. In the explicit report, we measured mean orientation adjustment bias, and in the implicit report, we measured reaction time in a visual search task for an oddly oriented line among heterogeneous distractors. Similar to other studies (Goldenberg et al., 2021; Kanaya et al., 2018), we observed a strong modulation of the explicitly reported mean by more salient items. However, the implicitly reported statistics were unaffected by saliency.

In Paper II, we implemented a procedure different from Kanaya et al. (2018), which allowed us to specify the properties of the amplification effect. First, we manipulated saliency via a task-irrelevant feature. Second, we employed the method of adjustment instead of 2AFC. In Kanaya et al. (2018), the amplification effect was proposed as a potential mechanism of non-random subsampling. Specifically, they hypothesized that the visual system can preferably sample a group of the most salient items, a number equal to the square root of the ensemble set size. Our results suggested that less salient items also contribute to the perceived mean. First, the observed bias toward the orientation of the most salient items was significantly smaller than expected if only the orientation of a sample of $\sqrt{N}$ most salient items was averaged. Second, the presence of less salient items also modulated the perceived mean. Specifically, the precision of participants' responses became higher when small items were added in the middle of the binomial orientation distribution of the large items. Conversely, when the small items were added at the extremes of the large item orientation distribution, the precision of averaging decreased.

5.3 Sampling

We examined the question of sampling in Papers I and II. In Paper I, we tested a computational model that implemented a neurally plausible mechanism of parallel exhaustive processing. In Paper II, we investigated whether only a subsample of the most salient items is integrated, or if all items are processed with further amplification. Importantly, here sampling is referring to information collection rather than the selection of whole items.

The results of Paper II suggest that both more salient and less salient items contribute to the perceived mean; however, the data do not allow us to draw strong conclusions regarding the number of averaged items. Generally, the observed contribution of less salient items and the larger weight of more salient items can be explained by three possible scenarios. First, the visual system can sample a small portion of items with a higher probability of more salient items being sampled. Then the disproportionate sample of more salient and less salient items is averaged with equal weights. The second scenario implies that all ensemble items are sampled; however, more salient items are weighted more highly during averaging. The third scenario is a combination of the previous two, meaning that more salient items are preferably sampled and have larger weights during averaging.

The results of Papers I and II together favor the scenario of parallel exhaustive processing with further weighting. The population coding model tested in Paper I suggested that all ensemble items are sampled, and the items' contribution to the perceived average is defined by the tuning curve width of the large receptive field neurons. The current version of the model does not account for focused attention modulation (e.g., an amplification effect), but it can be modified to do so (see Utochkin et al., 2024). Nevertheless, the findings of Papers I and II do not disprove the idea of subsampling but rather provide new arguments in favor of the parallel exhaustive processing theory.

5.4 Differences Between Explicit and Implicit Reports

Even though the population coding model can potentially explain the amplification effect for the perceived mean (Utochkin et al., 2024), the results of Paper III suggest that it cannot be directly applied to the representation of the whole feature distribution. The model utilizes the concept of the isomorphic neuron population response, which is similar to the concept of distribution shape representation even though these concepts are not directly related. The population coding model proposes that ensemble feature distributions are encoded in the form of the pooled neuron population response, which is isomorphic to the physical feature distribution. The reported ensemble average corresponds to the peak activation of this neuron response. Consequently, the systematic mean estimation bias towards more salient items would imply that the whole neuron population response distribution is skewed. This may lead to a hypothesis that item saliency would also affect the representation of the feature distribution shape. For example, a uniform orientation feature distribution would be encoded as skewed if items with more clockwise orientation relative to the mean orientation of the whole ensemble are more salient (e.g., bigger). Still, the results of Paper III showed no evidence supporting this hypothesis, suggesting that the mechanism of ensemble distribution shape encoding may differ from the one implemented in the population coding model.

Previous studies that have compared the properties of summary statistics reported explicitly and implicitly also reported fundamental differences between them. For example, Hansmann-Roth et al. (2021) suggested that observers are unable to explicitly distinguish ensembles with different feature distribution shapes, while implicit reports revealed that ensemble distribution shapes were encoded in rich detail. Moreover, explicit judgments of summary statistics have been shown to be more precise (Khayat et al., 2024), and the precision of explicitly and implicitly reported summary statistics was not correlated (Hansmann-Roth et al., 2021). Another difference concerns the number of simultaneously encoded ensembles. Studies suggest that observers can explicitly report summary statistics of two simultaneously presented ensembles without loss in precision (Chong & Treisman, 2005a) or with some decrease in task performance (Attarha et al., 2014; Attarha & Moore, 2015; Brand et al., 2012; Emmanouil & Treisman, 2008; Utochkin & Vostrikov, 2017). In contrast, the simultaneous implicit encoding of several ensemble statistics faces notable limitations, or it may even be impossible for observers in some cases (Hansmann-Roth et al., 2019). Taken together, there is growing evidence suggesting that explicit and implicit ensemble representations are implemented by partially non-overlapping mechanisms that may play different roles in perception. The results of Paper III revealed another difference between these types of reports: task-irrelevant salient features heavily influence explicit ensemble judgments, while implicitly reported statistics could be unaffected.

Still, a possible limitation of Paper III is the procedural difference in assessing implicit and explicit ensemble representations. To evaluate explicit representation, we employed the method of adjustment, exposing the ensemble only once for 500 ms. In contrast, the implicit representation was tested using the visual search task (i.e., the FDL paradigm; Chetverikov et al., 2016; 2019), where the ensemble was presented multiple times during prime trials, with each exposure lasting until the participant responded. Research indicates that the exposure duration can affect observers' performance in ensemble averaging tasks. In particular, the accuracy of explicit ensemble judgments increases with longer exposure durations, up to 500 ms (Whitney & Yamanashi Leib, 2018). Beyond this duration, further improvements in accuracy are not observed (Haberman et al., 2009; H. Li et al., 2016; Whiting & Oriet, 2011). Notably, the impact of exposure duration on amplification effects has not yet been studied conclusively.

6 Conclusions

Overall, our studies aimed to investigate sampling and weighting of information in ensemble perception. Specifically, we examined two types of weighting: weighting based on the items' deviation from the central tendency (i.e., robust averaging) and weighting based on item saliency (i.e., the amplification effect, Kanaya et al., 2018). To investigate item selection and robust averaging, we evaluated the population coding model (Utochkin et al., 2024) that utilizes parallel exhaustive item processing via neuronal population coding and robust averaging via synaptic weights (Paper I). To further explore the amplification effect, we modified the procedure of the original study of Kanaya et al. (2018) and implemented two types of reports: implicit (Paper II) and explicit (Paper III).

We found that the population coding model can accurately predict observers' behavior, suggesting that this model can be considered a neural implementation of parallel exhaustive processing and robust averaging in ensemble perception. Moreover, the model can be modified to account for the amplification effect studied in Paper II. In line with the previous study (Kanaya et al., 2018), in Paper II we found that the perceived mean is largely affected by more salient items using explicit reports. Additionally, we showed that less salient items also contribute to the perceived average, meaning that saliency modulates item contribution rather than absolutely determining their sampling. Still, in Paper III we observed no effect of saliency induced by a task-irrelevant feature on implicit ensemble representation, suggesting that the mechanism utilized by the population coding model cannot be directly applied to implicit ensemble representations. This difference between explicit and implicit ensemble representations aligns well with the hypothesis that they rely on separate mechanisms for their encoding.

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. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

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

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(18), 7345–7350. https://doi.org/10.1073/pnas.0808981106

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, *12*(2), 157–162. JSTOR.

Attarha, M., & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception, & Psychophysics*, *77*(4), 1116–1131. https://doi.org/10.3758/s13414-015-0870-0

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. https://doi.org/10.1037/a0036206

Baek, J., & Chong, S. C. (2020a). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics*, *82*(1), 63–79. https://doi.org/10.3758/s13414-019-01827-z

Baek, J., & Chong, S. C. (2020b). Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity. *Psychonomic Bulletin & Review*, 1–5. https://doi.org/10.3758/s13423-020-01718-7

Bauer, B. (2009). Does Stevens's Power Law for Brightness Extend to Perceptual Brightness Averaging? *The Psychological Record*, *59*(2), 171–185. https://doi.org/10.1007/BF03395657

Brand, J., Oriet, C., & Sykes Tottenham, L. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(1), 63. https://doi.org/10.1037/a0026950

Brezis, N., Bronfman, Z. Z., & Usher, M. (2017). A Perceptual-Like Population-Coding Mechanism of Approximate Numerical Averaging. *Neural Computation*, *30*(2), 428–446. https://doi.org/10.1162/neco_a_01037

Burr, D., & Ross, J. (2008). A Visual Sense of Number. *Current Biology*, *18*(6), 425–428. https://doi.org/10.1016/j.cub.2008.02.052

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, *153*, 196–210. https://doi.org/10.1016/j.cognition.2016.04.018

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017a). Rapid learning of visual ensembles. *Journal of Vision*, *17*(2), 21–21. https://doi.org/10.1167/17.2.21

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017b). Representing Color Ensembles. *Psychological Science*, *28*(10), 1510–1517. https://doi.org/10.1177/0956797617713787

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017c). Set size manipulations reveal the boundary conditions of perceptual ensemble learning. *Vision Research*, *140*, 144–156. https://doi.org/10.1016/j.visres.2017.08.003

Chetverikov, A., Hansmann-Roth, S., Tanrıkulu, Ö. D., & Kristjansson, A. (2020). Feature distribution learning (FDL): A new method for studying visual ensembles perception with priming of attention shifts. *Spatial learning and attention guidance*, 37-57. https://doi.org/10.1007/7657_2019_20

Chetverikov, A., & Kristjánsson, Á. (2024). Representing Variability: How Do We Process the Heterogeneity in the Visual Environment?. *Elements in Perception*. https://doi.org/10.1017/9781009396035

Choi, Y. M., & Chong, S. C. (2020). Effects of Selective Attention on Mean-Size Computation: Weighted Averaging and Perceptual Enlargement. *Psychological Science*, *31*(10), 1261–1271. https://doi.org/10.1177/0956797620943834

Chong, S. C., & Evans, K. K. (2011). Distributed versus focused attention (count vs estimate). *WIREs Cognitive Science*, *2*(6), 634–638. https://doi.org/10.1002/wcs.136

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

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. https://doi.org/10.3758/BF03195009

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. https://doi.org/10.1016/j.visres.2004.10.004

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *JOSA A*, 18(5), 1016-1026. https://doi.org/10.1016/j.visres.2005.07.037

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

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192. https://doi.org/10.1016/S0042-6989(97)00133-8

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*(5), 789–794. https://doi.org/10.3758/PP.70.5.789

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(Volume 18, 1995), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Elias, E., & Sweeny, T. D. (2020). Integration and segmentation conflict during ensemble coding of shape. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(6), 593–609. https://doi.org/10.1037/xhp0000733

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics*, *70*(6), 946–954. https://doi.org/10.3758/PP.70.6.946

Florey, J., Dakin, S., Clifford, C., & Mareschal, I. (2016). Summary Statistics for Gaze and Head Direction over Time. *Journal of Vision*, *16*(12), 499–499. https://doi.org/10.1167/16.12.499

Gardelle, V. de, & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, *108*(32), 13341–13346. https://doi.org/10.1073/pnas.1104517108

Goldenberg, A., Weisz, E., Sweeny, T. D., Cikara, M., & Gross, J. J. (2021). The Crowd-Emotion-Amplification Effect. *Psychological Science*. https://doi.org/10.1177/0956797620970561

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. Journal of Vision, 14(9), 22-22. https://doi.org/10.1167/14.9.22

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11), 1–1. https://doi.org/10.1167/9.11.1

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, *15*(4), 16. https://doi.org/10.1167/15.4.16

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753. https://doi.org/10.1016/j.cub.2007.06.039

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

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349).

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. https://doi.org/10.1038/nature07246

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple Spatially Overlapping Sets Can Be Enumerated in Parallel. *Psychological Science*, *17*(7), 572–576. https://doi.org/10.1111/j.1467-9280.2006.01746.x

Hansmann-Roth, S., Chetverikov, A., & Kristjánsson, Á. (2019). Representing color and orientation ensembles: Can observers learn multiple feature distributions? *Journal of Vision*, *19*(9), 2–2. https://doi.org/10.1167/19.9.2

Hansmann-Roth, S., Kristjánsson, Á., Whitney, D., & Chetverikov, A. (2021). Dissociating implicit and explicit ensemble representations reveals the limits of visual perception and the richness of behavior. *Scientific Reports*, *11*(1), 3899. https://doi.org/10.1038/s41598-021-83358-y

Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision*, *15*(4), 7–7. https://doi.org/10.1167/15.4.7

Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition*, *48*, 149–160. https://doi.org/10.1016/j.concog.2016.11.007

Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20172770. https://doi.org/10.1098/rspb.2017.2770

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of Directed Attention in the Human Extrastriate Cortex as Revealed by Functional MRI. *Science*, *282*(5386), 108–111. https://doi.org/10.1126/science.282.5386.108

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. https://doi.org/10.3758/s13414-023-02784-4

Khvostov, V. A., & Utochkin, I. S. (2019). Independent and parallel visual processing of ensemble statistics: Evidence from dual tasks. *Journal of Vision*, *19*(9), 3–3. https://doi.org/10.1167/19.9.3

Khvostov, V., Markov, Y., Brady, T., & Utochkin, I. (2020). *Limitations on animacy categorization in ensemble perception*. OSF. https://doi.org/10.31234/osf.io/d4za6

Kim, M., & Chong, S. C. (2020). The visual system does not compute a single mean but summarizes a distribution. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(9), 1013–1028. https://doi.org/10.1037/xhp0000804

Kristjánsson, Á., & Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research*, *48*(10), 1217–1232. https://doi.org/10.1016/j.visres.2008.02.007

Lee, H., Baek, J., & Chong, S. C. (2016). Perceived magnitude of visual displays: Area, numerosity, and mean size. *Journal of Vision*, *16*(3), 12–12. https://doi.org/10.1167/16.3.12

Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*, 13186. https://doi.org/10.1038/ncomms13186

Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of Individual Items during Ensemble Coding of Facial Expressions. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01332

Li, V., Castañón, S. H., Solomon, J. A., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. *PLOS Computational Biology*, *13*(8), e1005723. https://doi.org/10.1371/journal.pcbi.1005723

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, *15*(4), 6–6. https://doi.org/10.1167/15.4.6

Maule, J., & Franklin, A. (2016). Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism. *JOSA A*, *33*(3), A22–A29. https://doi.org/10.1364/JOSAA.33.000A22

Maunsell, J. H. R. (2015). Neuronal Mechanisms of Visual Attention. *Annual Review of Vision Science*, *1*(Volume 1, 2015), 373–391. https://doi.org/10.1146/annurev-vision-082114-035431

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences*, *111*(21), 7873–7878. https://doi.org/10.1073/pnas.1308674111

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11), 9–9. https://doi.org/10.1167/8.11.9

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

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.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*(1), 29–46. https://doi.org/10.1037/h0024722

Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision*, *13*(8). https://doi.org/10.1167/13.8.6

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, *11*(12), 18–18. https://doi.org/10.1167/11.12.18

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*(14), 19–19. 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), 13–13. https://doi.org/10.1167/11.12.13

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of Vision*, *18*(7), 4–4. https://doi.org/10.1167/18.7.4

Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze. *Psychological Science*, *25*(10), 1903–1913. https://doi.org/10.1177/0956797614544510

Teng, T., Li, S., & Zhang, H. (2021). The virtual loss function in the summary perception of motion and its limited adjustability. *Journal of Vision*, *21*(5), 2. https://doi.org/10.1167/jov.21.5.2

Utochkin, I. S., Choi, J., & Chong, S. C. (2024). A population response model of ensemble perception. *Psychological Review*, *131*(1), 36–57. https://doi.org/10.1037/rev0000426

Utochkin, I. S., & Vostrikov, K. O. (2017). The numerosity and mean size of multiple objects are perceived independently and in parallel. *PLoS ONE*, *12*(9). https://doi.org/10.1371/journal.pone.0185452

Watamaniuk, S. N. J., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*(5), 931–941. https://doi.org/10.1016/0042-6989(92)90036-I

Webb, B. S., Ledgeway, T., & McGraw, P. V. (2007). Cortical pooling algorithms for judging global motion direction. *Proceedings of the National Academy of Sciences*, *104*(9), 3532–3537. https://doi.org/10.1073/pnas.0611288104

Webb, B. S., Ledgeway, T., & McGraw, P. V. (2010). Relating spatial and temporal orientation pooling to population decoding solutions in human vision. *Vision Research*, *50*(22), 2274–2283. https://doi.org/10.1016/j.visres.2010.04.019

Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast! *Psychonomic Bulletin & Review*, *18*(3), 484–489. https://doi.org/10.3758/s13423-011-0071-3

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual Review of Psychology*, *69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Yamanashi Leib, A., Chang, K., Xia, Y., Peng, A., & Whitney, D. (2020). Fleeting impressions of economic value via summary statistical representations. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0000745

Original Publications

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals (I, II, III):

I. Iakovlev, A., & Utochkin, I. (2023). Ensemble averaging: What can we learn from skewed feature distributions?. Journal of Vision, 23(1), 5-5. https://doi.org/10.1167/jov.23.1.5

II. Iakovlev, A., & Utochkin, I. (2021). Roles of saliency and set size in ensemble averaging. Attention, Perception, & Psychophysics, 83, 1251-1262. https://doi.org/10.3758/s13414-020-02089-w

III. Iakovlev, A., Khvostov, V., Ásgeirsson, Á.G., Utochkin, I., Kristjánsson, Á. Amplification from saliency affects explicit but not implicit ensemble representations. Submitted for review in Attention, Perception, & Psychophysics. https://doi.org/10.31234/osf.io/yx4hr

All papers are reprinted by kind permission of the publishers.