Contextual Influences in Visual Perception:

Behavioural and Neural Insights into Size Representations

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Promotionsordnung vom 18.12.2018

vorgelegt von

Elif Memis

aus

Istanbul, Türkei

08 / 2025

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln am 05.02.2026 angenommen.

Dekanin oder Dekan: Prof. Dr. Jutta Stahl

1. Berichterstatterin oder Berichterstatter: Privatdozent Dr. Ralph Weidner
2. Berichterstatterin oder Berichterstatter: Jun.-Prof. Dr. Simone Vossel

Erklärung

Ich versichere eidesstattlich, dass ich die von mir vorgelegte Dissertation selbständig und ohne unzulässige Hilfe angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit einschließlich Tabellen, Karten und Abbildungen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe sowie dass diese Dissertation noch keinem anderen Fachbereich zur Prüfung vorgelegen hat. Die Promotionsordnung ist mir bekannt. Die von mir vorgelegte Dissertation ist von Jun.-Prof. Dr. Simone Vossel und Herrn Privatdozenten Dr. Ralph Weidner betreut worden.

Die Dissertationsschrift wurde von mir bicher weder im Inland noch im Ausland in gleciher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Juelich, den 05.02.2026

Author Contributions

This monograph-based dissertation with partial publications is based on the following original studies.

I. **Memis, E**., Yildiz, G. Y., Fink, G. R., & Weidner, R. (2025). Hidden size: Size representations in implicitly coded objects. Cognition, 256, 106041. https://doi.org/10.1016/j.cognition.2024.106041

The author of this dissertation: Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – original draft. *Yildiz, G.Y*: Visualization, Validation, Supervision, Software, Project administration, Methodology, Formal analysis, Conceptualization, Writing – review & editing. *Fink G.R*: Supervision, Resources, Methodology, Funding acquisition, Conceptualization, Writing – review & editing. *Weidner, R*: Validation, Supervision, Project administration, Methodology, Conceptualization, Writing – review & editing.

II. **Memis, E**., Fink, G. R., & Weidner, R. (*Unpublished manuscript, included in this dissertation and made publicly available via KUPS.*). Cross-ensemble size contrast in summary statistics: Neural and behavioural evidence.

The author of this dissertation: Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – original draft. *Fink G.R*: Supervision, Resources, Methodology, Funding acquisition, Conceptualization, Writing – review & editing. *Weidner, R*: Software, Formal analysis, Data curation, Validation, Supervision, Project administration, Methodology, Conceptualization, Writing – review & editing.

Table of Contents

1. Theoretical section 3

1.1. The visual system 3

1.1.1. Visual information flow through the central visual pathway 3

1.1.2. Hierarchical organisation of the visual cortex 6

1.1.3. Feedforward, feedback and recurrent processing 8

1.2. Mechanisms of size perception 10

1.2.1. Size perception and size constancy 10

1.2.2. Contextual modulation of size and rescaling 12

1.2.3. Ensemble summary statistics 14

1.2.4. Visual illusions 16

1.3. Visual information processing: from implicit mechanisms to predictive models 19

1.3.1. Implicit and explicit information processing in vision 19

1.3.2. Recurrent processing: insights from Object-Substitution Masking (OSM) 21

1.3.3. Predictive coding: a computational framework for visual perception 23

2. Empirical Section 26

2.1. Research Aims 26

2.2. Study I 28

2.2.1. Introduction 28

2.2.2. Methods 31

2.2.3. Experiment 1 - Screening: Ebbinghaus illusion 32

2.2.4. Experiment 2: Object substitution masking 39

2.2.5. Experiment 3: Size averaging 43

2.2.6. Experiment 4A: Size rescaling and masking 47

2.2.7. Experiment 4B: Size rescaling or size contrast? 54

2.2.8. Discussion 61

2.2.9. Conclusion 70

2.2.10. References 72

2.3. Study II 76

2.3.1. Introduction 76

2.3.2. Materials and methods 78

2.3.3. fMRI measurement 85

2.3.4. fMRI analysis 86

2.3.5. Behavioural results 92

2.3.6. fMRI results 97

2.3.7. Discussion 101

2.3.8. Conclusion 108

2.3.9. References 109

3. General Discussion 113

3.1. Implications 113
3.1.1. Visual illusions 113
3.1.2. Size perception 115
3.1.3. Contextual modulation in ensemble perception 117
3.2. Limitations and Future Directions 120
3.2.1. Study I 120
3.2.2. Study II 122
3.3. Concluding remarks 123
4. Summary 125
4.1. Study I 125
4.2. Study II 126
5. References 128
6. List of abbreviations 140
7. Acknowledgements 141
8. Curriculum Vitae 143

1. Theoretical section

Visual perception plays a central role in how we experience the world, enabling us to identify objects, navigate spaces, and interpret social cues. However, given the brain's limited processing capacity, the visual system employs various strategies to efficiently extract meaningful information from complex environments. One such strategy is creating ensemble summary statistics, allowing observers to rapidly and accurately extract the global properties of a scene, such as the average size of a group of objects. Yet, size perception is not simply a passive registration of the external world. Instead, it is modulated by contextual influences, as nicely illustrated by size contrast illusions, such as the Ebbinghaus illusion. This thesis explores the extent to which contextual modulation influences ensemble representations, and whether these representations themselves serve as context-inducing standards in size judgments.

Study I examines whether objects that are implicitly coded due to object-substitution masking (OSM) still contribute to ensemble representations in a size-rescaled manner. Study II investigates whether different sets of stimuli, such as task-relevant and task-irrelevant objects differentiated by colour, interact in ways that produce mutual size contrast effects. Together, these investigations aim to clarify whether summary statistics extend beyond mere descriptions of object features and instead actively influence the perceived size of objects.

This chapter includes the theoretical foundation for the thesis, organised into three key sections: (1.1) The visual system, (1.2) Mechanisms of size perception, and (1.3) Visual information processing: from implicit mechanisms to predictive models. Each section explores the mechanisms that collectively inform our understanding of how the brain constructs and modulates size representations.

1.1. The visual system

1.1.1. Visual information flow through the central visual pathway

The visual system processes raw sensory input in a structured sequence, ultimately producing meaningful visual representations of the external world. This system converts the incoming sensory input into neural signals that travel along the central visual pathway to the visual cortex (Fig. 1.1.1-1). The process begins when the eye captures light, and the first anatomical structure the light encounters is the cornea (Schaeffel, 2006). It is located at the front of the eye and plays an important role in adjusting focus by bending the incoming light. From there, the light moves through the pupil and reaches the lens, which is an adjustable structure responsible for the focus. Specifically, the lens adapts its curvature based on the distance of the object being viewed. For example, the lens has a more curved shape when focusing on a nearby object, while it flattens for objects that are farther away. This process is called accommodation, and ensures that objects are sharply projected on the retina's photoreceptor layer.

Next, light reaches the retina, which has multiple layers and includes numerous photoreceptor cells that respond to light (Kolb, 2003). These photoreceptor cells contain protein molecules called opsins. When opsins absorb photons, they trigger a chemical reaction inside the photoreceptor. This reaction changes the cell's electrical state and causes hyperpolarisation. There are two types of photoreceptors, each with different contributions to visual processing. Rods are sensitive to low levels of light, thus facilitating vision under dark conditions. In contrast, cones respond to high levels of light and are responsible for processing colour. The allocation of rods and cones varies across the retina. Specifically, rods are more concentrated in the peripheral retina, while cones are densely located in the foveal region. Output signals generated by rods and cones are transmitted to bipolar cells, and then to retinal ganglion cells. Later, the optic nerves are formed when all ganglion cell axons converge. The optic nerves are essential, as they allow the transmission of visual information from the eye to the brain (Boycott & Wässle, 1999).

Once the signals leave the retina, the optic nerves from both eyes connect at the optic chiasm. This is the point where axons are reorganised into two optic tracts based on which side of the visual field they originated from. In detail, signals originating from the inner (nasal) part of each retina are routed to the opposite hemisphere, while those from the outer (temporal) remain on the ipsilateral side. This process is known as contralateral projection, ensuring that some sensory input from the left visual field is also processed in the right hemisphere, and some input from the right visual field is directed to the left hemisphere.

Later, the reorganised signals continue along the optic tracts and arrive at the lateral geniculate nucleus (LGN). This structure includes six distinct layers, and each layer receives input from either the contralateral or ipsilateral eye (Merigan & Maunsell, 1993; Nassi & Callaway, 2009). For example, input from the contralateral eye is handled by layers 1, 4 and 6, whereas layers 2, 3 and 5 take input from the ipsilateral eye (Wurtz & Kandel, 2000). The LGN consists of magnocellular and parvocellular layers. The magnocellular neurons process motion and are sensitive to alterations in brightness but not to colour. Neurons in these layers have larger receptive fields and respond to low spatial frequencies. In contrast, the parvocellular neurons are specialised in detecting fine details and colour. They have smaller receptive fields and respond to high spatial frequencies. Parvocellular neurons transmit signals more slowly than magnocellular neurons. Finally, the processed visual signals are projected from the LGN to the primary visual cortex for more detailed processing.

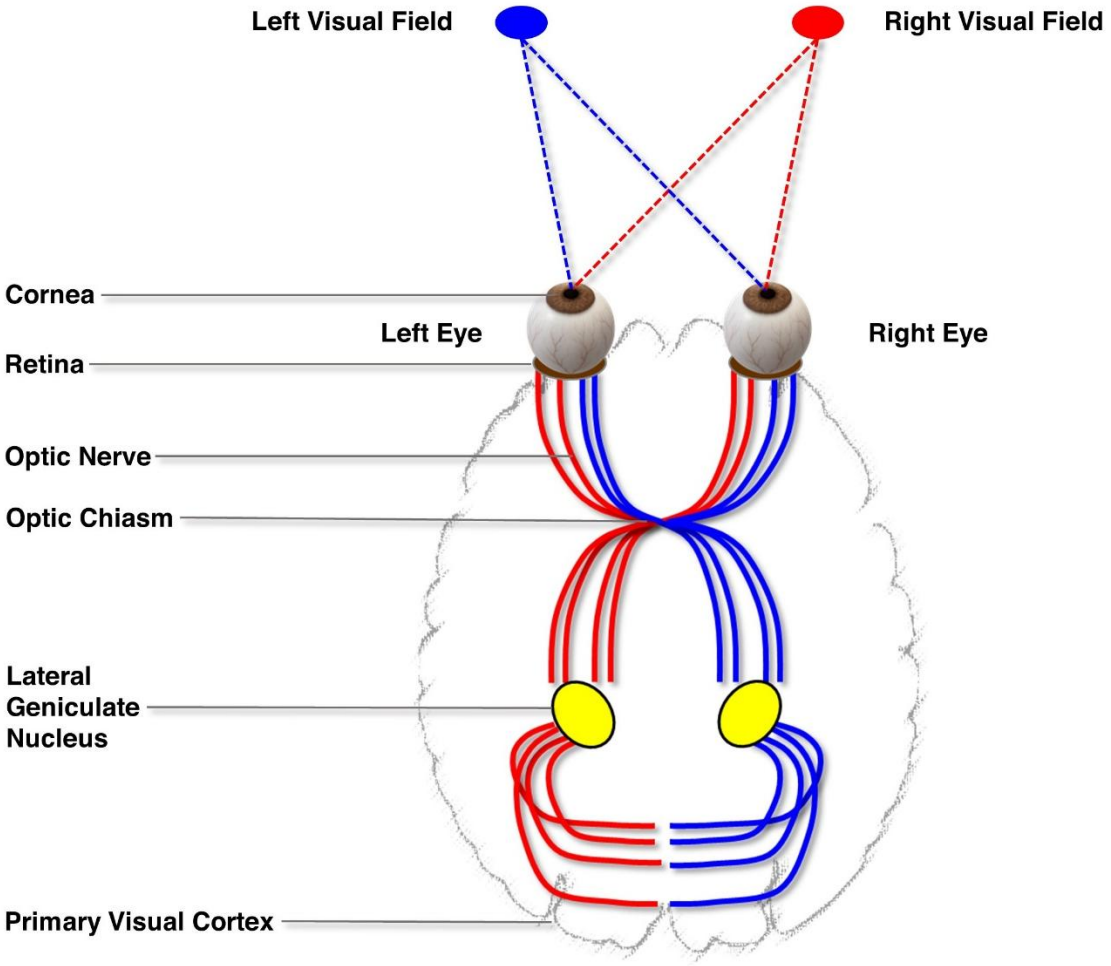

Figure 1.1.1-1 *Illustration of information processing in the visual system, including the cornea, eye, retina, optic nerve, optic chiasm, LGN, and primary visual cortex.*

1.1.2. Hierarchical organisation of the visual cortex

Visual information arriving from the eyes through the central visual pathway reaches the visual cortex, located in the occipital lobe (Fig. 1.1.2-1). The first region receiving direct input from the LGN is the primary visual cortex (V1), also known as the striate cortex, because of its distinct layered structure (Callaway, 1998). This region serves as the first stage where sensory signals are processed. V1 is divided into six distinct layers, each with a specific role in handling the incoming visual input. The input arriving from the retina via the

LGN enters layer 4, which is primarily composed of simple cells responsible for the initial processing of sensory input (Hubel & Wiesel, 1962; Callaway, 1998; Anderson & Martin, 2009). Layers 2 and 3 receive strong feedforward projections from layer 4 and integrate this information before sending outputs to higher visual areas. Layer 1 is the most superficial one, and receives modulatory feedback from higher-order visual areas, whereas complex cells are mostly found in layers 2, 3, and 5. After the visual information is processed in V1, V2 processes more complex aspects of the visual scene, such as changes in angles, shape and object orientation (Anzai, Peng & Van Essen, 2007).

Beyond V2, the visual system contains several specialised extrastriate areas, including V3, V4, and MT (also known as V5). Specifically, V3 is involved in processing dynamic form and shape integration (Felleman & Van Essen, 1991). V4 is responsible for processing object recognition and colour perception, whereas MT focuses on analysing motion direction and speed (Felleman & Van Essen, 1991; Zeki, 1993; Born & Bradley, 2005). These extrastriate areas are functionally divided into two processing pathways (Fig. 1.1.2-1): the ventral stream and the dorsal stream (Mishkin & Ungerleider, 1982; Goodale & Milner, 1992). The ventral stream originates in V1 and extends to the temporal lobe, referred to as the "what" pathway and is responsible for object identification and recognition. In parallel, the dorsal stream projects toward the parietal cortex and is referred to as the "where" or "how" pathway, supporting the spatial processing and the visual guidance of actions. While this division was originally based on their distinct roles in object recognition and spatial localisation, subsequent studies have shown that the ventral and dorsal streams are also differentially involved in conscious and unconscious visual processing. The dorsal stream mediates visually guided actions without conscious awareness of the visual stimuli, whereas the ventral stream is closely linked to conscious recognition of objects (Goodale & Milner, 1992; Goodale, Westwood & Milner, 2004).

Moreover, the visual system is retinotopically organised, meaning that neighbouring locations in the retina are represented at neighbouring locations on the cortical surface (Kaas, 1997; Wandell, Dumoulin, & Brewer, 2007). Retinotopic maps are most clearly defined in early visual areas, whereas this spatial organisation becomes less pronounced in higher-order visual areas, due to larger receptive fields and the emergence of more complex organisational principles. Reflecting this retinotopic organisation, Study II employed a position localizer task to define distinct quadrants in early visual areas, confirming that stimuli presented in specific locations elicited spatially distinct cortical activations consistent with the retinotopic organisation of the visual cortex.

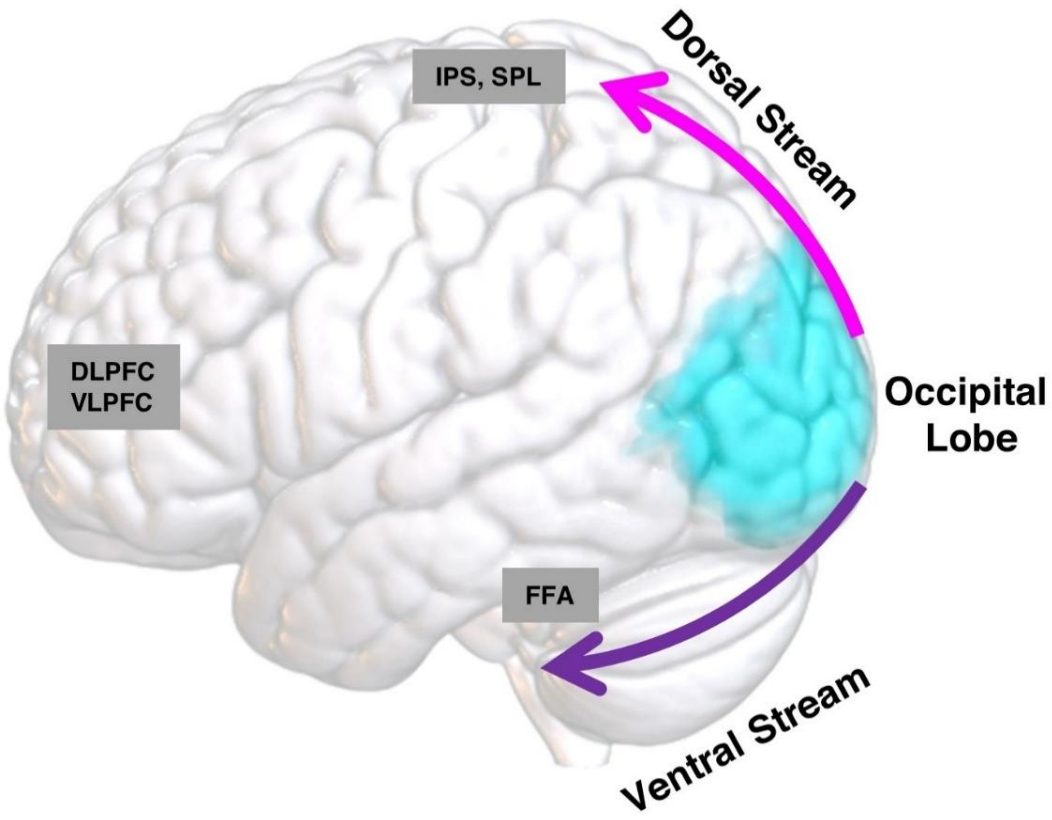

Figure 1.1.2-1 *Illustration of the dorsal and ventral stream.*

1.1.3. Feedforward, feedback and recurrent processing

The information is hierarchically processed in the visual cortex. Early visual areas, such as V1, are specialised for detecting basic features like edges and orientations, while

higher-order areas represent increasingly complex properties such as object detection and motion (Hubel & Wiesel, 1962; Felleman & Van Essen, 1991; Grill-Spector & Malach, 2004). This hierarchical fashion is accompanied by an enlargement of the receptive field size and an increase in the complexity of encoded features. One fundamental model of information processing in the visual system is the feedforward sweep, which is a stimulus-driven approach that reflects a bottom-up flow of information. Namely, the information is initially transmitted from the retina through the LGN, and then processed hierarchically along successive cortical stages. During the processing, each stage is responsible for extracting and integrating more complex aspects of the stimuli.

Although the feedforward sweep reflects the initial processing of visual information, it alone cannot account for the context-sensitive and predictive nature of visual perception. This model assumes that information processing relies entirely on incoming sensory input without any top-down influence from higher-level regions. If perception operated only in this stimulus-driven fashion, it would lack adaptability and fail to incorporate prior knowledge.

To account for this, the feedback mechanism, often referred to as top-down processing, has been proposed (Kanwisher & Wojciulik, 2000). In this framework, information processing is expectation-driven and involves signals from higher-level cortical regions to project back to earlier processing stages. These top-down signals allow early visual areas to be modulated based on context, expectations and prior knowledge, indicating that perception is not a passive registration of the external world.

Previous studies have contributed to the distinction between feedforward and feedback mechanisms, showing that the initial feedforward sweep of information predominantly supports pre-attentive processing, while feedback modulations are essential for attentive tasks and visual awareness (Lamme & Roelfsema, 2000). For example, the top-down influence of figure-ground segregation has been suggested to involve feedback to

striate cortex from higher-order visual areas (Zipser, Lamme & Schiller, 1996). Similarly, evidence from a TMS study has demonstrated that motion awareness requires top-down feedback from MT to the early visual cortex (Pascual-Leone & Walsh, 2001).

However, increasing evidence emphasises the importance of recurrent processing, which suggests that perception does not rely on a one-way flow of information (Lamme & Roelfsema, 2000). Instead, it emerges from iterative interactions between feedforward and feedback signals across cortical levels. This mechanism enables the visual system to generate perceptual hypotheses at higher levels and test them against the signals present in early visual areas (Di Lollo et al., 2000; see Chapter 1.3.2).

1.2. Mechanisms of size perception

1.2.1. Size perception and size constancy

Precisely perceiving an object's size is essential for visually guided actions. Whether reaching for a nearby item or estimating the dimensions of objects at a distance, the visual system must transform the retinal input into a stable real-world representation. An important aspect of visual perception is known as size constancy, which refers to the ability to perceive objects as maintaining a constant size despite changes in the retinal image due to variations in viewing distance (Boring, 1940; Holway & Boring, 1941). For example, when an object moves closer, its image becomes larger on the retina, yet we don't experience that object as physically growing. Instead, we perceive it as maintaining a constant size. This perceptual constancy in various circumstances reflects the brain's ability to incorporate depth cues to interpret changing retinal signals in a stable manner.

Exploring the neural basis of size constancy requires investigating how object size is represented across multiple brain regions involved in visual information processing. This processing follows a hierarchical progression through specific brain regions (see Chapter 1.1.2). In the early stages, the primary visual cortex initially processes the incoming visual

input transmitted from the LGN. As processing moves up the hierarchy, areas within the ventral and dorsal streams develop increasingly complex representations of object properties and spatial relationships. For instance, Konkle and Caramazza (2013) investigated whether object size (e.g., large vs. small) and animacy (e.g., animals vs. objects) elicit distinct neural response patterns. Results revealed that real-world object size acts as a main organising principle in the ventral pathway. Specifically, medial regions, including the parahippocampal cortex, responded more strongly to large objects, whereas the inferior temporal gyrus showed greater activation for small objects. Interestingly, this size-based cortical organisation was found only for inanimate objects, and similar effects were not observed when participants were shown animate objects. Supporting this view, another study using multivoxel pattern analysis (MVPA) revealed that real-world object size plays a crucial role in structuring object representations across the human visual cortex (Julian et al., 2017). They found that object size not only influenced overall activation strength but also shaped the distribution of response patterns in specific brain regions. Notably, regions associated with scene perception, such as the parahippocampal place area (PPA), the retrosplenial complex (RSC), and the occipital place area (OPA), showed stronger activations to large objects. In contrast, the occipitotemporal sulcus (OTS) responded more strongly to small objects, consistent with findings from other studies (Konkle & Oliva, 2012; Konkle & Caramazza, 2013).

Crucially, the perceived size of an object is influenced not only by its retinal image but also by various factors such as depth cues, surrounding objects, and prior knowledge. Murray and colleagues (2006) were among the first to demonstrate that perceived size, rather than retinal size, modulates neural activity in the primary visual cortex (V1). In the experiment, participants viewed two identical spheres placed within a depth-inducing Ponzo corridor. Although the spheres had identical retinal sizes, the one positioned higher in the corridor was perceived as farther and, therefore, larger. This perceptually larger sphere

elicited a significantly broader spatial activation in V1 compared to the one perceived as smaller. These findings suggest that early visual areas not only encode retinal object size but are also influenced by top-down contextual feedback. Building on these findings, Fang and colleagues (2008) replicated the results by demonstrating that V1 encodes the perceived size of objects, rather than their physical size. They also found that activity in higher-level areas, such as the lateral occipital cortex (LOC) and the PPA, was significantly reduced during a demanding fixation task. This suggested that narrowed attentional focus disrupted top-down feedback from higher cortical areas, which is essential for integrating depth cues. Further supporting these findings, Sperandio and colleagues (2012) showed that the extent of cortical activation in V1 reflected the perceived, rather than retinal, size of the afterimage.

1.2.2. Contextual modulation of size and rescaling

Size perception requires more than simply reading the raw retinal image. While retinal size provides an initial sensory input, the visual system integrates a range of factors, such as contextual cues, depth and distance, to form the perceived size of objects. As retinal size varies with distance, the visual system must adjust or rescale the perceived size accordingly. This size rescaling refers to the perceptual adjustment of an object's size influenced by contextual influences (Gilinsky, 1951; Kilpatrick & Ittelson, 1953). A foundational framework explaining this mechanism is the size-distance invariance hypothesis, which proposes that perceived size (S') depends on both the retinal size (α) and the perceived distance (D'). This relationship is expressed as:

$$S' = 2D' \tan(\alpha/2),$$

indicating how the visual system combines retinal input with perceived distance to compute the perceived object size across varying viewing conditions. Recent studies have advanced our understanding of size rescaling mechanisms, particularly in the context of ensemble representations–the ability to extract statistical properties like average size from

groups of objects (Ariely, 2001). For example, Tiurina and Utochkin (2019) investigated whether average size judgments are computed based on the retinal or perceived size of objects. In the experiment, participants were shown a set of objects at different distances and asked to report the average size. They found that average size judgments reflect the perceived, not the retinal, average size of the items, indicating that rescaling occurred prior to averaging. Based on their findings, the visual system appears to first incorporate depth cues to bind size information, and only then compute the average size. Extending these findings, Markov and Tiurina (2021) examined whether rescaling mechanisms influence not only mean size estimation but also range estimation within an ensemble. By using both binocular and monocular cues, they demonstrated that observers estimated the average range of object sizes only after size-distance rescaling occurred. Together, these findings indicate that contextual rescaling precedes both mean and range estimations in size perception.

In addition to behavioural studies highlighting how perceived object size is rescaled through contextual information, neuroimaging studies provide further insight into the underlying neural dynamics of this modulation. For example, Zeng and colleagues (2020) used TMS to disrupt activity in both the early visual cortex (EVC) and LOC during a size discrimination task, including a hallway background. Disruption in either region reduced the strength of the illusion, indicating that both areas contribute to contextual modulation of size. Crucially, the effect of TMS in LOC occurred earlier than in EVC, suggesting that size information is first integrated into higher visual areas and then projected back to early visual regions to modulate low-level representations. Consistent with this, previous studies demonstrated a broader network of higher-level visual areas in illusory size perception, including the LOC (Weidner & Fink, 2007), right superior parietal lobule and right superior parietal cortex (Plewan et al., 2012), as well as prefrontal regions (Libedinsky & Livingstone, 2011; Schall, 2015). Together, these findings are consistent with the idea that size

representations might be shaped by interactions between feedforward sensory input and top-down feedback, potentially through recurrent processing.

1.2.3. Ensemble summary statistics

Visual scenes contain far more information than the visual system can entirely process, yet humans navigate complex environments with remarkable efficiency. One key strategy underlying this competency is the extraction of ensemble summary statistics, which refers to the ability to compute summary statistics from groups of objects rather than processing each item individually (Ariely, 2001). This mechanism allows observers to rapidly perceive global properties such as average size, orientation, and density. Ariely (2001) conducted pioneering experiments on summary statistics, demonstrating that observers accurately estimate the average size of a set of circles, despite being unable to report the individual sizes of the objects. Subsequent studies reinforced this idea across a wide range of visual features, including orientation (Dakin & Watt, 1997), speed (Watamaniuk & Duchon, 1992), emotional expressions (Haberman & Whitney, 2007), motion direction (Sweeny, Haroz & Whitney, 2013), and body size (Oswald, 2023). Interestingly, Chong and Treisman (2005) showed that the ability to extract ensemble statistics is not restricted to one set of stimuli. They found that average size estimates for two simultaneously presented sets were as precise as those for a single set, indicating that multiple ensemble representations can be maintained in parallel. These findings provide strong evidence that ensemble summary statistics are computed rapidly, automatically, and effortlessly.

Several theoretical frameworks have been proposed to explain how ensemble summaries are computed. According to the summary-statistic model, most or all items are processed in parallel to extract global properties (Ariely, 2001; Chong & Treisman, 2005). In contrast, the subsampling hypothesis suggests that ensemble statistics are formed by averaging a restricted number of randomly chosen items, instead of analysing the entire set

(Myczek & Simons, 2008). Meanwhile, robust or weighted averaging models propose that observers assign different weights to individual elements based on their reliability or consistency (de Gardelle & Summerfield, 2011). These models support the idea that ensemble summaries are not merely simple arithmetic means, but rather weighted estimates in which less reliable items, such as outliers, contribute less. Additionally, global interaction models propose that ensemble perception may emerge through nonlinear computations performed across all items within the visual field (Jia et al., 2022). These frameworks fundamentally differ in their assumptions about processing capacity, attentional modulation, and the computational stage at which statistical summaries emerge.

Ensemble perception is not solely determined by the physical properties of the stimuli but is influenced by contextual factors (see Chapter 1.2.2). For example, Im and Chong (2009) investigated whether mean size estimates are based on the physical or perceived average size of the objects by manipulating the context through Ebbinghaus inducers. Their findings showed that mean size judgments were modulated by the size of the Ebbinghaus inducers, indicating that size averaging is based on the perceived rather than the physical size of the to-be-averaged set. In addition to contextual modulation, another critical question concerns the level of processing at which ensemble representations emerge. Choo and Franconeri (2010) addressed this by examining whether average size estimates can be computed when the visibility of objects is impaired by OSM, a type of visual masking thought to disrupt object representations at later stages of processing while preserving initial encoding (see Chapter 1.3.2). The authors hypothesised that if ensemble representations are formed before the stage at which OSM interferes, then both visible and invisible items would contribute to the average size estimates. Conversely, if masked items fail to influence the average, this would imply that ensemble coding depends on later-level object representations. They found that both visible and invisible objects contributed to mean size judgments,

providing strong evidence that ensemble statistics can be computed at an early stage of visual processing. Building on these findings, study I combined the contextual influences and the level of processing within the same paradigm to investigate whether objects whose visibility is impaired by OSM still contribute to average size in a rescaled manner through Ebbinghaus inducers.

Behavioural evidence strongly supports the view that ensemble summary statistics can be accessed rapidly, even before individual item features are fully processed. Consistent with this, an EEG study demonstrated that neural signals related to ensemble properties emerge rapidly in visual processing prior to individual item representations (Epstein & Emmanouil, 2021). Meanwhile, other studies highlight the contribution of higher-level cortical regions to ensemble coding. Using a frequency-tagging EEG approach, Jia and colleagues (2022) isolated neural responses to systematically varying ensemble sizes, discovering that signals in parieto-occipital areas tracked global ensemble size but not individual item sizes. Additional support from fMRI studies showed that ensemble representations are encoded in the anterior medial ventral visual cortex, particularly in the PPA, as revealed through an fMRI adaptation paradigm (Cant & Xu, 2012). In a follow-up study, Cant and Xu (2015) extended these findings by examining what types of ensemble features are represented in this area. They found that the PPA was sensitive to changes in relative density, but not to changes in spacing between objects. This suggests that PPA encodes higher-level ensemble properties rather than low-level spatial features. In contrast, LOC responded to changes at the local level, such as the shape of objects, indicating that LOC processes object-level features within ensembles.

1.2.4. Visual illusions

Visual illusions demonstrate that perception is not a direct reflection of the external world, but rather an active process shaped by contextual information and prior knowledge. Rather than indicating perceptual failures, illusions reflect the brain’s adaptive mechanism to

integrate environmental cues. A widely recognised theory proposed by Gregory (1963, 1968) suggests that some visual illusions arise from the brain's effort to maintain size constancy (see Chapter 1.2.1). When depth cues lead to misleading interpretations of distance, this internal rescaling mechanism adjusts perceived size inappropriately, making objects appear larger or smaller than they actually are. For instance, in the Ponzo illusion, two horizontal lines of identical length are placed over converging lines, resembling a linear perspective cue, such as a railway track, which induces depth information (Fig. 1.2.4-1). The upper line appears longer because it is perceived as being farther away, leading the brain to rescale its size to maintain size constancy; thus, this line is perceptually rescaled as longer (Ponzo, 1911; Leibowitz et al., 1969). Similarly, the moon illusion shows that the moon is perceived as noticeably larger when it is close to the horizon than when it is high in the sky, despite having the same retinal size in both cases (Ross & Plug, 2002). This effect is thought to result from the presence of depth cues along the horizon, which make the moon appear farther away, hence, it is rescaled as larger by the visual system (Kaufman & Kaufman, 2000; Weidner et al., 2014).

Moreover, the Ebbinghaus illusion shows that size perception is shaped by contextual factors even in the absence of depth cues, as identical objects appear to differ in size depending on the size of the surrounding items (Ebbinghaus, 1902). In detail, the target circle appears smaller when surrounded by large inducers and larger when surrounded by small inducers. Traditionally, earlier accounts associated this illusion with size contrast effects (Massaro & Anderson, 1971), proposing that surrounding circles serve as standards or benchmarks for size judgements and have a contrast-like effect on perceived size. Conversely, other theories emphasised low-level contour interactions (Chen et al., 2018; Todorović & Jovanović, 2018). Supporting this viewpoint, it has been demonstrated that the magnitude of the illusion strengthens as the similarity between the target and inducer

increases (Rose & Bressan, 2002). An interesting study using continuous flash suppression and backward masking (Chen et al., 2018) showed that the Ebbinghaus illusion can still be elicited even when either the target or the inducers are suppressed from conscious awareness. Therefore, the illusory effect induced by the Ebbinghaus illusion appears to occur during early visual processing. Similarly, it has been shown that individuals with smaller V1 areas experience stronger Ebbinghaus illusion, indicating that anatomical variability in the early visual cortex is a predictor for illusion strength (Schwarzkopf, Song, & Rees, 2011).

Furthermore, comparing different illusions has become a useful approach for understanding how the visual system constructs perceptual interpretations, and whether these processes rely on shared or distinct neural mechanisms. For example, Song and colleagues (2011) discovered the effect of interocular transference on the Ebbinghaus and the Ponzo illusion. They found that the strength of the Ponzo illusion remained comparable whether shown to the same or different eyes, whereas the strength of the Ebbinghaus illusion diminished. These results align with evidence that the Ebbinghaus illusion depends on lateral interactions within the primary visual cortex (Bosking et al., 1997; Schwarzkopf, Song, & Rees, 2011), whereas the Ponzo illusion seems to involve top-down feedback from higher-level regions responsible for interpreting a three-dimensional spatial context (Fang et al., 2008). Likewise, Wu and colleagues (2023) extended this view by showing that although both illusions engage the right superior parietal lobule (SPL), the Ebbinghaus illusion relies on feedforward signals from V1 to SPL, but the Ponzo illusion is shaped by internal dynamics within the right SPL itself. While there is strong evidence for early sensory origin in the Ebbinghaus illusion, a recent study using fMRI and DCM revealed contributions from higher-level cortical areas (Chen et al., 2024). They found enhanced feedback connectivity from the right precuneus to the extrastriate cortex during size overestimation.

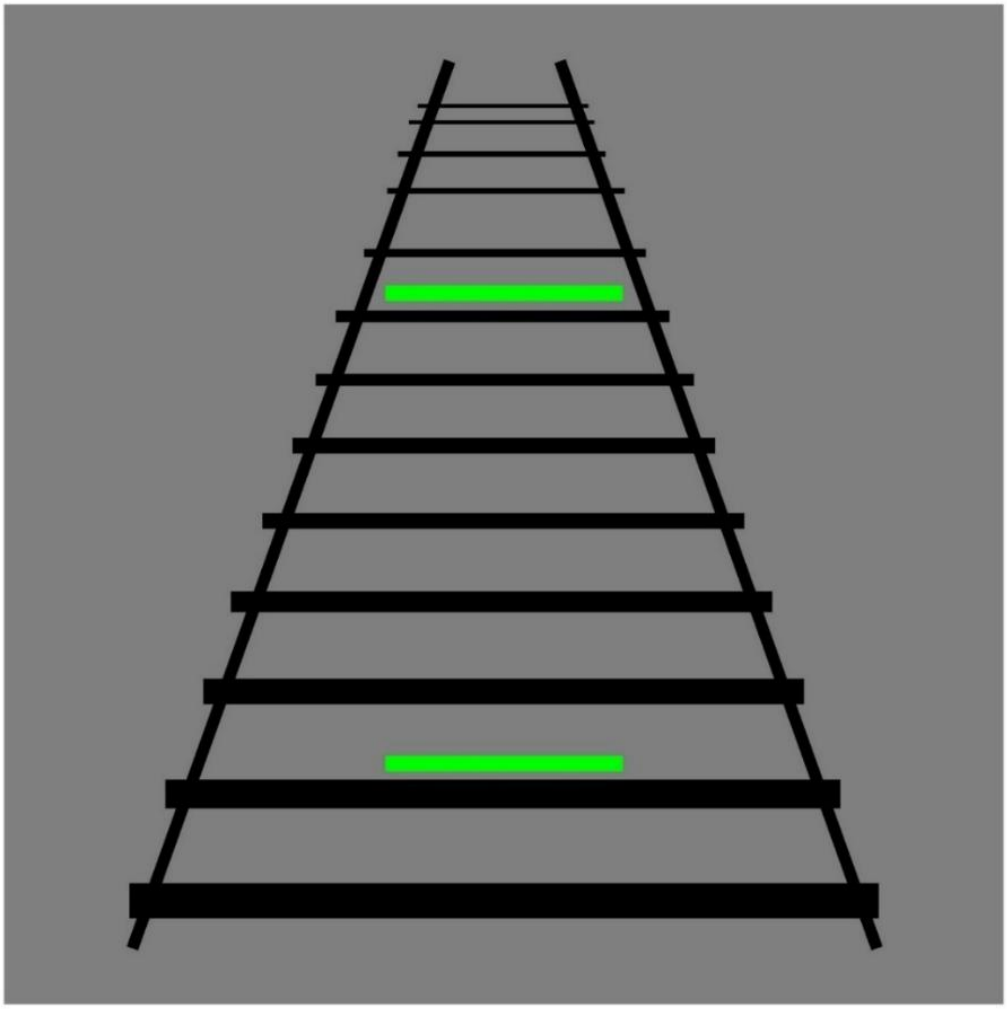

Figure 1.2.4-1 *Illustration of the Ponzo illusion.*

1.3. Visual information processing: from implicit mechanisms to predictive models

1.3.1. Implicit and explicit information processing in vision

Although each input is transmitted from the central visual pathway to the visual cortex, not all of them are explicitly reportable by an observer. The distinction between implicit and explicit information processing has gained significance following the work of Graf and Schacter (1985) on memory. They defined implicit memory as a performance advantage that occurs without conscious recollection, while explicit memory is characterised by the necessity of conscious retrieval of prior experiences. In the domain of visual perception, this dichotomy has been extended to distinguish between explicit visual

processing, which involves the conscious perception and reportability of stimuli, and implicit visual processing, in which visual information is registered and influences behaviour despite the absence of conscious awareness (Kouider & Dehaene, 2007). Findings from studies on blindsight (Aleci & Dutto, 2024), subliminal priming (Fischer et al., 2024) and object recognition (Tal et al., 2024) suggest that the brain can process visual input even when it is not explicitly reportable.

To explain the neural mechanisms behind these phenomena, researchers have introduced models such as the global neural workspace (GNW) and the recurrent processing framework. The GNW model suggests that consciousness arises when sensory information is widely distributed throughout the brain, particularly to frontoparietal regions, thereby making it globally accessible to multiple areas (Dehaene & Changeux, 2011). In contrast, the recurrent model states that consciousness emerges through feedback connections, allowing information to be processed in recurrent loops without necessarily requiring global distribution to frontal regions (Lamme, 2003).

Building on these models, several studies have explored the contributions of the dorsal and ventral visual streams at different processing levels. For example, Fang and He (2005) investigated this by employing interocular suppression to render images (e.g., tools and faces) invisible. Their findings revealed that the dorsal pathway responds to invisible images, and this effect was observed when participants were shown tools instead of faces. Meanwhile, other studies have revealed that the ventral stream is more closely associated when the stimuli are explicitly reportable by observers (Tong et al., 1998; Bar et al., 2001).

One of the most robust demonstrations of implicit visual processing is the phenomenon of blindsight, also known as residual vision. In this condition, individuals with damage to the primary visual cortex (V1) can detect and respond to stimuli in their blind visual field even though they cannot explicitly report the presence of the stimuli (Aleci &

Dutto, 2024). Their ability to correctly guess features, such as motion and orientation, indicates that visual information is still being processed outside of awareness. Neuroimaging studies have supported this notion by showing preserved activity in extrastriate areas, such as the human motion area hMT+, LOC, intraparietal sulcus (IPS), and precuneus (Ajina & Bridge, 2019; MacLean et al., 2023).

Beyond clinical populations, attention plays a crucial role in implicit and explicit information processing. Inattentional blindness (IB), famously demonstrated by the invisible gorilla experiment (Simons & Chabris, 1999), refers to the failure to notice a visible, unexpected stimulus because attention is occupied elsewhere. IB nicely demonstrates how stimuli can be present in the visual field yet remain unreportable when unattended (Hutchinson et al., 2019). For example, an interesting study tested whether perceptual inference occurs for stimuli that are not explicitly reportable by using the IB paradigm with fMRI (Vandenbroucke et al., 2014). During the task, participants were given a demanding fixation task to maintain their attention focused on the task, while they were presented peripherally with either Kanizsa or control figures to create the IB effect. Interestingly, results showed enhanced activation in the visual areas V1, V2, V3, V3ab, V4, and LOC in response to the Kanizsa figure, despite the figures being coded implicitly. Furthermore, MVPA analysis revealed that responses to invisible Kanizsa figures reliably predicted responses to visible figures, particularly in LOC and V3ab, suggesting that perceptual inference doesn’t necessarily require the explicit coding of the stimuli.

1.3.2. Recurrent processing: insights from Object-Substitution Masking (OSM)

OSM is an effective masking paradigm to investigate the temporal dynamics of visual information processing. When a target object appears briefly with a four-dot mask that persists beyond the target's offset, target identification becomes significantly impaired. Notably, the effectiveness of OSM is robust when specific conditions are met, such as

distributed attention, short target exposure (< 100 ms), delayed mask duration compared to the target, and loss of visual awareness of the target (Goodhew et al., 2013). OSM is considered an ideal candidate for studying recurrent processing because OSM disrupts late object representation while preserving initial feedforward sensory input (Di Lollo et al., 2000). The prolonged presence of the mask leads to additional feedforward processing of the mask alone, which interferes with the recurrent processing needed to consolidate the target-mask representation. Consequently, the target fails to reach conscious awareness, even though early visual encoding remains intact. Compared to other masking techniques, such as pattern masking and metacontrast masking, which disrupt early visual processing (Enns, 2004; Bugmann & Taylor, 2005), OSM, therefore, provides strong evidence that conscious perception depends not merely on the arrival of visual input, but on successful recurrent interactions within the cortical hierarchy.

An interesting study by Choo and Franconeri (2010) investigated whether size averaging depends on early sensory representations or on later stages of processing. In order to test this, they employed OSM and hypothesised that if size averaging relies on early representations, performance would remain comparable even when items are masked, whereas if it depends on later processing, masked items would not contribute to the ensemble. Their results showed that both visible and invisible objects influenced size judgments, suggesting that size averaging occurs at an early stage of visual processing, and does not require fully consolidated object representations to shape perception. In Study I, we extend these findings by showing that the contribution of masked objects to the ensemble is already rescaled through Ebbinghaus inducers.

Neurophysiological evidence further supports that OSM offers an important experimental tool for distinguishing rapid sensory encoding from later stages of information processing. For example, previous research indicates that the N170 component, which

reflects object recognition, showed a decrease in OSM (Reiss & Hoffman, 2007). Furthermore, another study by Carlson and colleagues employed the functional magnetic resonance adaptation (fMR-A) technique to assess how repeated neural responses were affected by OSM (Carlson, Rauschenberger & Verstraten, 2007). While adaptation effects were observed in early visual areas such as V1, they were diminished in higher-order regions like the LOC, indicating that OSM interferes with later stages of processing while preserving the initial feedforward input. Additionally, Weidner and colleagues (2006) have identified a cortical network for OSM, including early visual areas (e.g., V1), the middle occipital gyrus, the transverse occipital gyrus, and the left IPS (Weidner, Shah & Fink, 2006). These regions showed the strongest activation during effective masking, when the visibility of the target was highly impaired. The authors suggested that these findings indicate the presence of a cortical network involved in initiating and iteratively testing perceptual hypotheses.

1.3.3. Predictive coding: a computational framework for visual perception

While recurrent processing refers to iterative interactions between higher-level and lower-level cortical regions, the predictive coding framework extends this notion by proposing that such feedback connections carry learned expectations derived from past experiences (Rao & Ballard, 1999). According to this framework, the brain operates as a Bayesian inference system, continuously checking top-down predictions with incoming sensory data. For that reason, predictable information is suppressed to minimise redundancy and enable efficient processing by allocating resources to novel or unexpected stimuli. When there is a mismatch between the predicted and the incoming input, called the prediction error, only this error signal is transmitted forward through the hierarchy.

One of the key assumptions of predictive coding theory is expectation suppression, in which predicted stimuli elicit diminished neural responses. The reason behind this effect has been interpreted in multiple ways, including as a form of filtering (Rao & Ballard, 1999), as

redundancy reduction (Murray, Schrater & Kersten, 2004), or by the sharpening of neural representation, where expected features are enhanced and unexpected activity suppressed (Kok et al., 2012). For example, Rao and Ballard's (1999) predictive coding model accounts for end-stopping, a receptive-field phenomenon where a neuron's firing rate decreases when a stimulus extends beyond its classical receptive field. This reduction in response is interpreted as a sign of efficient neural processing, where predicted input leads to decreased activity. In this framework, higher-level regions send predictions about the expected activity to lower-level sensory areas, while the lower-level areas return only the residual error, which is the discrepancy between the actual sensory input and the prediction.

Within the predictive coding framework, attention is one of the most influential factors in modulating neural activity. Kok and colleagues (2012) investigated the interaction between attention and prediction in shaping neural activations using an orientation discrimination task in fMRI. During the task, attention was manipulated via presenting spatial cues, whereas prediction was modulated by varying the probability of the stimuli. They found that neural activity in V1 decreased when the stimuli were predicted and unattended, supporting the notion of expectation suppression. However, this suppression was absent when the stimuli were attended, and instead, they observed an increased activation in V1. These results support the predictive coding account in which attention enhances the precision of prediction errors, thereby making the brain more sensitive to unexpected or task-relevant input.

Additionally, the interaction between attention and prediction has been investigated through a Bayesian computational framework (Feldman & Friston, 2010). They modelled a classical Posner spatial cueing paradigm, where performance advantage occurs when the cue is presented at expected locations. Their model suggests that attention helps the brain determine how much to trust incoming sensory signals by modulating the influence of

prediction errors. When a signal is considered reliable, attention increases its impact on perception. This helps in explaining why we respond faster and more accurately to stimuli that are both expected and attended. Overall, these findings suggest that attention modulates perception by increasing the precision of prediction errors (Feldman & Friston, 2010).

Beyond functional imaging and computational models, neurophysiological evidence further supports the predictive coding theory, as evidenced by an ERP known as mismatch negativity (MMN) (Garrido et al., 2009). MMN reflects the brain's reaction to unexpected events, and it is observed when there is a mismatch between the prediction and the actual input. It originates primarily in the auditory cortex but is also influenced by higher-order cortical regions. Although MMN was first identified in the auditory domain, similar mismatch responses have also been found in the visual system (Czigler et al., 2004), suggesting that prediction error is a general mechanism operating across sensory modalities. For example, Stefanics and colleagues (2018) used EEG and computational modelling to test whether visual mismatch negativity (vMMN) reflects trial-by-trial prediction errors. In their study, participants viewed a series of face images that changed in colour or emotional expression. The researchers used a Bayesian model to simulate how the brain might predict upcoming stimuli and how strongly it reacts when those predictions are violated. They found that brain activity was better explained by trial-wise prediction errors than by a change detection model. Overall, predictive coding provides a strong framework for understanding perception as an active, expectation-driven process, in which the brain continuously anticipates, evaluates, and updates its interpretation of sensory input.

2. Empirical Section

2.1. Research Aims

This chapter presents behavioural and fMRI studies investigating how the visual system extracts ensemble summary statistics. These studies focus on how contextual factors shape ensemble representations and whether these representations themselves act as benchmarks for size judgments.

I. Study I investigated the levels of processing at which ensemble summary statistics are formed by examining whether implicitly coded objects contribute to perceived average size. To manipulate the perceived average size, three out of eight circles were surrounded by Ebbinghaus inducers. OSM was implemented by keeping the inducers on the screen longer than the target circles. The purpose of employing that particular type of masking was to disrupt late-level object representations while preserving early sensory encoding, allowing us to assess whether the influence of the inducers on the ensemble calculation occurred before or after masking. If context integration on the ensemble occurs prior to masking, then even masked objects would influence the perceived average size, resulting in either an increase or a decrease, depending on the inducer size. Conversely, if contextual effects alter size representations solely after the explicit recognition of an object, then the masked items would not be rescaled by the inducers, and no modulation of perceived average size would be expected. This study, therefore, examined whether implicitly coded objects are integrated into ensemble statistics in a size-rescaled manner.

II. In a control experiment testing the general effect of Ebbinghaus inducers in Study I, we observed that simultaneous presentation of two sets of stimuli is not coded independently, but rather influences the size judgments of the task-relevant items.

Building on this finding, Study II examined whether the simultaneous presentation of two sets of stimuli influences their perceived size in a contrast-like fashion, and whether such effects are reflected in altered functional activation patterns within retinotopically defined visual areas. Although behavioural responses were limited to the task-relevant set, our quadrant-specific design made it possible to examine whether the task-irrelevant set was also subject to size contrast effects at the neural level. Specifically, we investigated whether increasing the average size of task-irrelevant items would result in a decrease of the perceived size of task-relevant items. If such a size contrast effect exists, the perceived average size of the task-relevant set would be altered by the size of the task-irrelevant set, and this modulation would also be reflected in corresponding changes in functional activation patterns. Alternatively, if there is no such effect, we would expect no significant modulation in either behavioural responses or functional activations across the experimental conditions. To disentangle these possibilities, the study employed a quadrant-specific design to evaluate both behavioural responses and functional markers associated with size contrast.

2.2. Study I

***Memis, E**., Yildiz, G. Y., Fink, G. R., & Weidner, R. (2025). Hidden size: Size representations in implicitly coded objects. Cognition, 256, 106041. https://doi.org/10.1016/j.cognition.2024.106041*

2.2.1. Introduction

Perceiving our environment is an overwhelming computational challenge for our visual system, given that the processing capacities of our perceptual system are limited. At the same time, the amount of information in the outside world is infinite. One strategy the brain uses to cope with the vast amount of information entering our visual system via the retina is to process ensemble representations rather than individual items in the visual scene (Ariely, 2001; Chong & Treisman, 2005). The visual system forms ensemble representations by averaging certain visual features (i.e., size, orientation, speed) of items belonging to the same perceptual group (Ariely, 2001; Alvarez & Oliva, 2008; Watamaniuk & Duchon, 1992) and hence can efficiently code the average size of multiple objects. Interestingly, these representations only indirectly contain individual item features, so while observers can compare the average size of groups of items, they struggle to report visual features of individuated items in these groups (Ariely, 2001).

A recent study demonstrated that the perceived average size of a group of items changes based on size-distance rescaling mechanisms, indicating that the average estimates were computed after size-distance rescaling (Markov & Tiurina, 2021). Thus, the ensemble representation contributing to the formation of an average size of multiple objects is not merely a low-level coding of the angular size represented on the retina. Instead, it is a size value that resembles size as we perceive it. This is particularly intriguing in light of previous findings indicating that the brain estimates the average size of a group of items by considering the size of items that are not consciously perceived (Choo & Franconeri, 2010). These findings suggest that it should be possible to assess the size representations of masked

and, hence, more implicitly coded objects and test whether context information modifies their size representations.

To answer the question of whether size averaging involves the retinal size of masked and, hence, more implicitly coded objects or if it is based on perceived size, i.e., involving size representations that are already size-rescaled, we combined object-substitution masking (OSM) (Enns & Di Lollo, 1997) with the Ebbinghaus illusion. The paradigms were linked so that the stimuli constituting the mask also served as inducers generating the Ebbinghaus illusion.

In OSM, an object is masked if the surrounding dots persist on the screen longer than the object itself. OSM is ideal for studying implicitly coded objects since it is assumed to alter only later representations while keeping early representations intact (Di Lollo et al., 2000). Changing the size of these dots allows the generation of Ebbinghaus-like inducer patterns. In the classic Ebbinghaus illusion, a target surrounded by small Ebbinghaus inducers appears larger than its physical size, while a target surrounded by large Ebbinghaus inducers appears smaller than its physical size (Ebbinghaus, 1902; Massaro & Anderson, 1971). To investigate early contextual modulations of size, it is important to use an illusion that operates at early levels of visual processing. It is known that the Ebbinghaus illusion is weaker when the inducer and target stimuli are presented to separate eyes, indicating the involvement of monocular neurons in early visual processing (Song et al., 2011; Nakashima & Sugita, 2018). This is also supported by fMRI studies, which indicate an important role for V1 in the Ebbinghaus illusion (Schwarzkopf & Rees, 2013; Schwarzkopf et al., 2011). The Ebbinghaus illusion is therefore particularly well suited as a means of altering early size representations. In this study, six surrounding dots were presented to operate as Ebbinghaus inducers and induce OSM. This way, we could manipulate both the perceived size of the target objects and their level of encoding.

To establish that each of these paradigms generates the expected effects, three behavioural experiments were conducted before the main experiment, testing whether the stimulus configurations used in our experiments a) allow to induce changes in perceived sizes using our Ebbinghaus stimulus configuration (Experiment 1), b) allow to generate a robust masking effect using OSM (Experiment 2) and c) allow to confirm that size averaging was sensitive enough to detect size changes in the magnitude of the ones induced by our Ebbinghaus stimulus configuration (Experiment 3). Having successfully tested these three phenomena in isolation, we combined the Ebbinghaus illusion, OSM and size averaging into one paradigm (Experiment 4A-B) to investigate whether masked and, therefore, less recognisable objects are already size-rescaled. Participants were presented with a typical size-averaging paradigm, while six red inducers surrounded three out of eight green circles to induce both the Ebbinghaus illusion and OSM.

We hypothesised that the size of the inducers would generate an Ebbinghaus effect on the target objects, leading to a change in their perceived size, either an increase or a decrease. Additionally, because the target objects were included in the group of items used for the size averaging task, we expected that the altered size representations would also influence the perceived average size, as previously demonstrated by Markov and Tiurina (2021). The current study's experimental design also permits investigating the role of the recognisability of the target stimulus and its contribution to perceived size averaging. In particular, if context integration via the Ebbinghaus illusion alters size representations before an object is recognised, then the size of even masked objects would incorporate size-averaging mechanisms in a size-rescaled manner. If this is the case, participants would report a larger perceived average size in blocks with small inducers and a smaller perceived size in blocks with large inducers, regardless of whether the targets are masked. This pattern would be

consistent with size-scaling effects that operate at an early level of processing, even before a stimulus is recognised.

Alternatively, if masked objects are not yet affected by context effects and, therefore, unchanged by the Ebbinghaus illusion, we would expect no difference between large and small inducer blocks.

2.2.2. Methods

Overview of experiments

To investigate whether context-induced size-rescaling affects representations of masked objects, we combined OSM with the Ebbinghaus illusion in a size-averaging paradigm. In a series of experiments, we tested each of these paradigms separately to confirm the efficiency of the respective experimental manipulation.

General Setup

All experiments were presented via Visual Studio 1.68.1 using PsychoPy 2021.2.3 scripts on an AORUS F048U 47.53-inch monitor at a distance of 57 cm. The screen resolution was 1920 x 1080 pixels with a refresh rate of 120 Hz. The distance between participants and the monitor was maintained using a chin- and forehead rest. Each participant completed the experiment in a darkened room, and self-paced breaks were presented between the experimental blocks. All four experiments involve the presentation of a circular array (10 degrees of visual angle) of equally spaced eight green (76.71 cd/m^2) circles around a black (0.00 cd/m^2) fixation cross. Red inducers with a luminance of 18.27 cd/m^2 were used in all experiments, except for Experiment 3. The distance between the large inducers and the targets was 1.5 degrees of visual angle and the distance between the small inducers and the target was 1.2 degrees of visual angle. The distance between adjacent green circles was 3.8 degrees of visual angle. Additionally, the distance from the centre to the green circles is 5

degrees of visual angle, and the distance from the centre to the nearest inducer was 3.5 degrees of visual angle. The distances between the stimuli are comparable to those in similar studies (Choo & Franconeri, 2010; Jacoby et al., 2013). Experimental stimuli were presented over a grey (49.90 cd/m^2) background at 40 degrees of visual angle over a black screen. Pre-cues (used in Experiment 1) and post-cues (used in Experiment 2) shared the same luminance value as the green circles (76.71 cd/m^2).

2.2.3. Experiment 1 - Screening: Ebbinghaus illusion

Experiment 1 had two primary objectives. First, we aimed to test whether our stimulus configurations reliably altered a stimulus's perceived size by surrounding it with either large or small inducers. Second, we aimed to identify participants exhibiting a strong illusion effect, who we recruited for the subsequent experiments.

Participants

Sixty healthy participants (32 females) attended Experiment 1. To identify participants with a strong illusion, the criterion for inclusion in subsequent experiments required participants to exhibit an illusion strength of at least ten per cent (>10%) in small (0.9 degrees of visual angle) target conditions of Experiment 1 (see Stimuli subsection for details). A 10% illusion effect is comparable to typical values observed in the previous studies (Chen et al., 2024; Wu et al., 2023, Chen et al., 2018). This criterion applied to a group of twenty-nine participants (M = 29.38 years, SD = 5.52, 17 females) who participated in the remaining experiments on separate days. All participants had normal or corrected-to-normal vision. We obtained written informed consent before the experiment and paid all participants 10 euros per hour for their participation. The ethics committee of the German Society of Psychology approved the study. The sample size was determined based on effect sizes detected in previous studies (Choo & Franconeri, 2010) to ensure the current sample size would be sufficient to reveal significant

differences between the experimental conditions (in a repeated-measures analysis of variance) with 95% power and an alpha level of 0.05. Power estimates were computed using G*Power (Erdfelder et al., 1996).

Stimuli

One out of eight circles was randomly selected as a target circle, surrounded by either large (Fig. 1A) or small (Fig. 1B) circular inducers, generating an Ebbinghaus figure. We used the same stimulus configuration but without Ebbinghaus inducers as a control condition (Fig. 1C). Since there were no inducers in the control condition to locate the target circle, a pre-cue (Fig. 1, time window 2) was employed in all experimental conditions to indicate the target position. The target size was either 0.9 (small target) or 1.1 (large target) degrees of visual angle. Incorporating two different target sizes allowed us to test whether participants correctly performed the task.

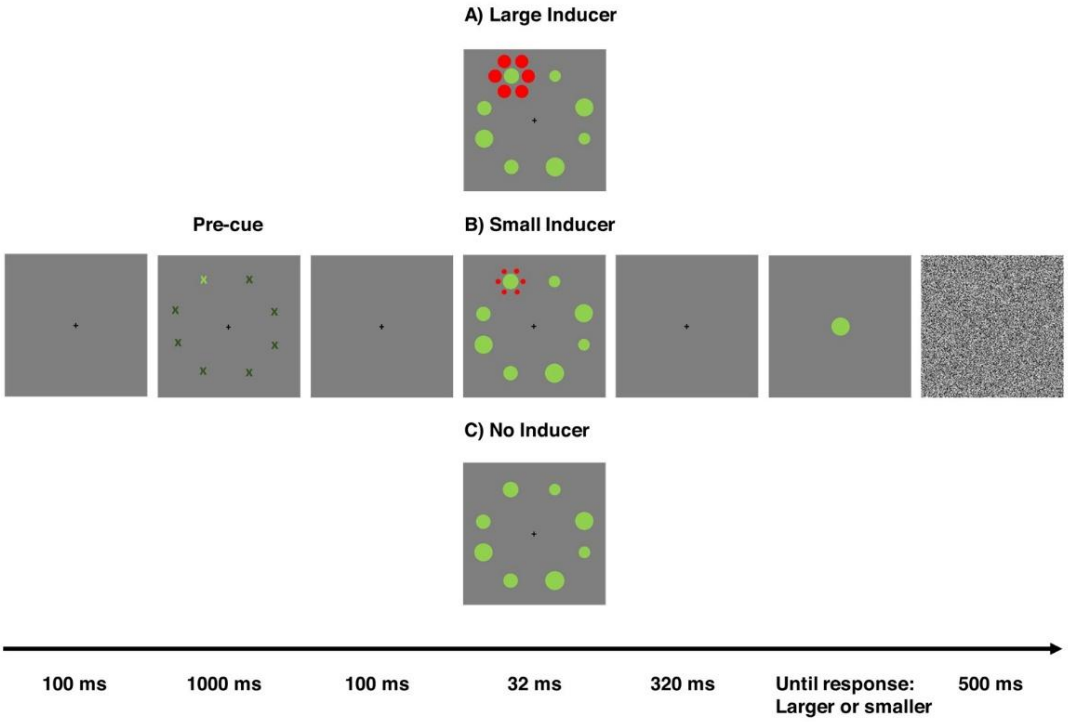

Fig. 1. Illustration of the task in Experiment 1. *Participants were asked to determine whether the comparison circle (time window 6) was larger or smaller than the target circle (time window 4). The position of the target circle was indicated by a pre-cue (time window 2). The type of inducer was manipulated as either a large inducer (A) or a small inducer (B), while no inducers (C) were presented in the control condition.*

The small and large inducers were always presented at 0.7 and 1.3 degrees of visual angle, respectively. The average size of the seven distractor circles was determined in accordance with the target size to keep the physical average size of the screen equal across different target sizes. In detail, the distractor average size was 0.9 degrees of visual angle for the small target and 1.1 degrees of visual angle for the large target. The average size of the distractors was calculated based on the normal cumulative distribution in all experiments. Specifically, the size of each circle was varied with a standard deviation of 0.15 degrees of visual angle around the mean (0.9 or 1.1 degrees of visual angle). This resulted in distinct

sizes for each distractor circle in each trial while keeping the mean size of the seven distractors constant.

The Method of Constant Stimuli was used to detect the perceived size of the target circle. The comparison circle's size varied around the target stimulus's with .1 degrees of visual angle increments, resulting in two different comparison size lists. Ten different comparison sizes were used (half of them were smaller than the size of the target stimulus, and the other half was larger). Each comparison circle was presented ten times.

Procedure

Participants were presented with a size judgment task where each trial started with a 100 ms fixation in which only the fixation cross was presented over a light grey background, followed by a pre-cue for 1000 ms (Fig. 1). The pre-cue was used to indicate the position of the target circle. After a fixation presentation for 100 ms, green circles appeared around the fixation cross in a circular array. Participants saw red inducers around the pre-cued green circle in the experimental blocks with the small and large inducers. All stimuli disappeared simultaneously after 32 ms, and only the fixation cross was displayed on the screen for 320 ms. A green comparison circle was then displayed at the centre of the screen until participants pressed a button to indicate whether the comparison circle was larger or smaller than the target circle (indicated by a pre-cue). Each trial ended with a grey noise pattern for 500 ms. Participants completed 600 trials (2 target sizes × 3 inducer types × 10 comparison sizes × 10 repetitions), resulting in an experiment length of around 40 min.

Statistical Analyses

All statistical analyses were performed using the JASP software package version 0.16.0.0 (University of Amsterdam, Amsterdam, Netherlands). We measured the perceived size of the target stimulus in Experiment 1. To assess how participants perceived the target

stimulus's size, we generated psychometric curves for each participant across six experimental conditions. These curves were constructed by analysing response proportions at precise intervals of 0.1 degrees, showing a preference for the comparison circle being larger than the target circle. We employed the logistic function to quantify the probability (*P*) of perceiving the comparison circle as larger than the target circle. Subsequently, the Point of Subjective Equality (PSE) was calculated as $P = 0.5$, representing the point where the comparison circle was perceived at an equivalent size to the target circle. Width of each psychometric curve was calculated as $P_{0.75} - P_{0.5}$, representing the degree of uncertainty in the participants' responses for each experimental condition. Higher values of curve width indicate a greater perceptual uncertainty. We calculated goodness of fit measures when fitting psychometric curves to the data. The obtained curves demonstrated a strong fit in Experiment 1 (*r* ranged between .728 and .989).

The PSEs are anticipated to shift towards the right in the case of small inducer conditions, indicating that the target size is perceived as larger than its physical size. Conversely, we would anticipate an opposite shift towards the left in large inducer conditions, indicating that the target size is perceived as smaller than its physical size.

In Experiment 1, the obtained PSEs and curve widths were entered into separate 2 x 3 ANOVAs with the factors target size (small, large) and inducer type (small inducer, large inducer, no inducer) to test whether the Ebbinghaus illusion altered the stimulus size.

Results

Figure 2 displays the mean PSEs with standard errors for within-subject contrast for the small (Fig. 2A) and large (Fig. 2B) target sizes, along with different inducer-type conditions. ANOVA revealed significant main effects for both target size [small vs. large] ($F(1, 28) = 111.81$, $p < .001$, $\eta^2_p = 0.800$) and inducer type [small inducer vs. large inducer vs.

no inducer] (F (2, 56) = 46.35, p < .001, η^2_p = 0.623). Moreover, the interaction between target size and inducer type was also significant (F (2, 56) = 5.86, p = .005, η^2_p = 0.173).

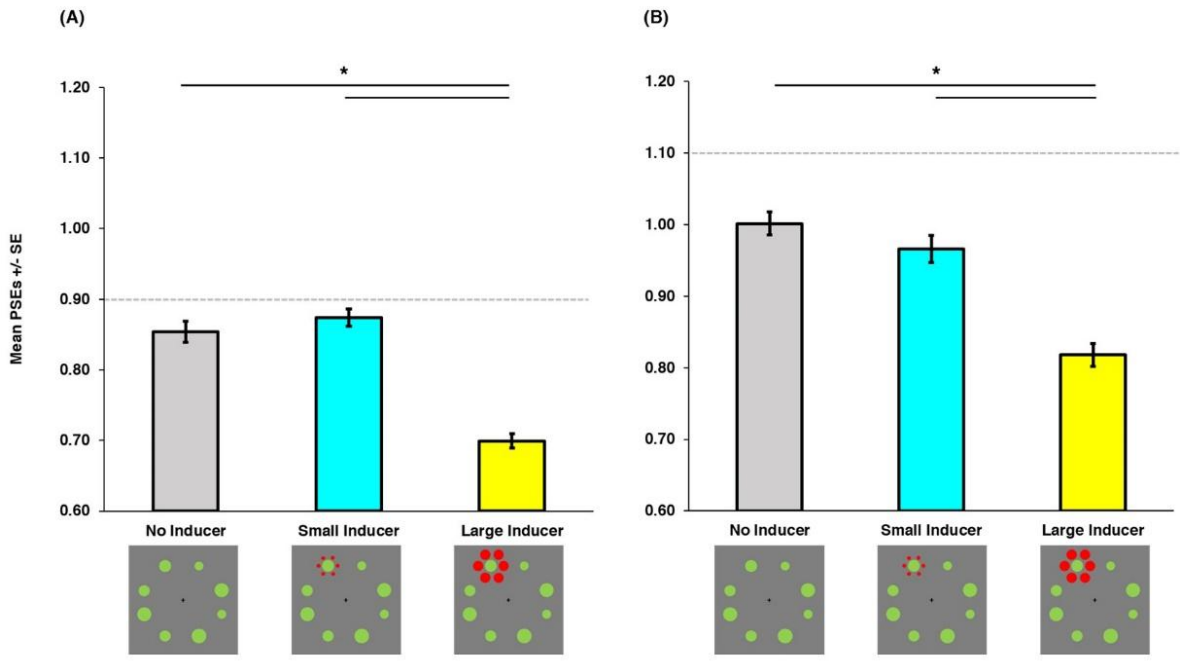

Fig. 2. Perceived size of the target stimulus for small (A) and large (B) target sizes in Experiment 1. *Averaged PSEs across different inducer types were plotted. Grey bars indicate the no inducer condition, blue bars represent the small inducer condition and yellow bars represent the large inducer condition. Asterisks (*) represent significant differences at p < .05. Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014). The horizontal dashed grey lines represent the physical size of the target stimulus. The figures shown below the x-axis are illustrations of the corresponding experimental conditions.*

Above all, participants estimated the perceived size of the target circle as larger in the small inducer conditions (M = 0.92, SE = 0.03) than in the large inducer conditions (M = 0.76, SE = 0.03) [planned t-test, t (28) = 8.40, *Bonferroni corrected* p < .001, Cohen's d = 1.560], indicating a robust illusory size judgment (Fig. 2A-B). Furthermore, target size

estimation was greater in the no inducer conditions (M = 0.93, SE = 0.02) compared to the large inducer conditions (M = 0.76, SE = 0.03) [planned t-test, t (28) = -8.96, *Bonferroni corrected* $p < .001$, Cohen's d = -1.664]. However, the mean PSEs did not show a significant difference between the conditions with small inducers and those without inducers for any target size [planned t-test, t (28) = 0.59, *Bonferroni corrected* p = 1.870, Cohen's d = 0.109].

Consistent with our expectations, the averaged PSEs were higher in the large target condition (M = 0.93, SE = 0.03) than in the small target condition (M = 0.81, SE = 0.02). The inducer type and target size interaction suggests that the effect of inducer type on the mean PSEs varies based on target size. Overall, these findings indicate that inducer type significantly impacted participants' size judgment (illusion effect for small target: 19.12%), showing that our variant of the Ebbinghaus illusion significantly altered the perceived size of our target stimulus.

Curve widths for the small and large targets were compared across inducer types (small inducer, large inducer, no inducer) by conducting a 2 x 3 ANOVA. ANOVA showed a significant main effect for inducer type [small inducer vs. large inducer vs. no inducer] (F (2, 55) = 3.256, p = .047, Greenhouse-Geisser corrected, η^2_p = 0.104). Neither the main effect of target size (p = .137) nor the interaction between target size and inducer type was significant (p = .282). Specifically, the estimated curve widths were significantly greater in the small inducer conditions (M = 0.21, SE = 0.02) compared to the no inducer conditions (M = 0.17, SE = 0.01) [planned t-test, t (28) = -2.50, *Bonferroni corrected* p = .046, Cohen's d = -0.419]. However, the estimated curve widths did not reveal a significant difference between the large inducer conditions and the no inducer conditions (p = .286), and between the large inducer conditions and the small inducer conditions ($p \geq .999$).

2.2.4. Experiment 2: Object substitution masking

Since our primary goal was to investigate hidden size representations, we employed OSM to reduce target recognisability. Specifically, we tested whether the stimuli used as Ebbinghaus-like inducers in Experiment 1 could efficiently induce OSM by manipulating their presentation times. In a simultaneous viewing condition, the mask and the green circles disappeared simultaneously after 32 ms, aiming to preserve their accessibility levels to conscious visual awareness. In the delayed viewing condition, the mask persisted on the screen 320 ms longer than the green circles, inducing a masking effect.

Method

Participants

Experiment 2 featured a group of twenty-nine participants (M = 29.38 years, SD = 5.52, 17 females), all exhibiting an illusion effect of >10 per cent in Experiment 1.

Stimuli

Eight green circles were presented around the fixation cross, and two circles, randomly chosen (one from the right visual field, the other one from the left), were surrounded by either large (Fig. 3A) or small (Fig. 3B) inducers. This diagonal arrangement of inducers was intended to ensure a diffuse distribution of attention during the entire task. One of the green circles surrounded by red inducers was randomly defined as the target stimulus, while the other was assigned as a nontarget stimulus. Target and nontarget stimuli were presented at either 0.8 or 1.1 degrees of visual angle. The target and nontarget sizes were independent from each other. Trials where the target and nontarget stimuli were the same size were categorized as congruent, while those with different-sized target and nontarget stimuli were classified as incongruent. To prevent unequal distributions, we ensured that target and nontarget sizes had the same size in half of the trials (i.e., congruent trials: both target and

nontarget at 0.8 or 1.1 degrees of visual angle), while in the remaining trials, they were different (i.e., incongruent trials: target stimulus at 0.8 degrees of visual angle and nontarget stimulus at 1.1 degrees of visual angle and vice versa). The participant was informed which of the stimuli marked by inducers was task-relevant using a post-cue. The average size of the six distractor circles was always 0.95 degrees of visual angle, and the average size was calculated based on the normal cumulative distribution (M = 0.95, SD = 0.15).

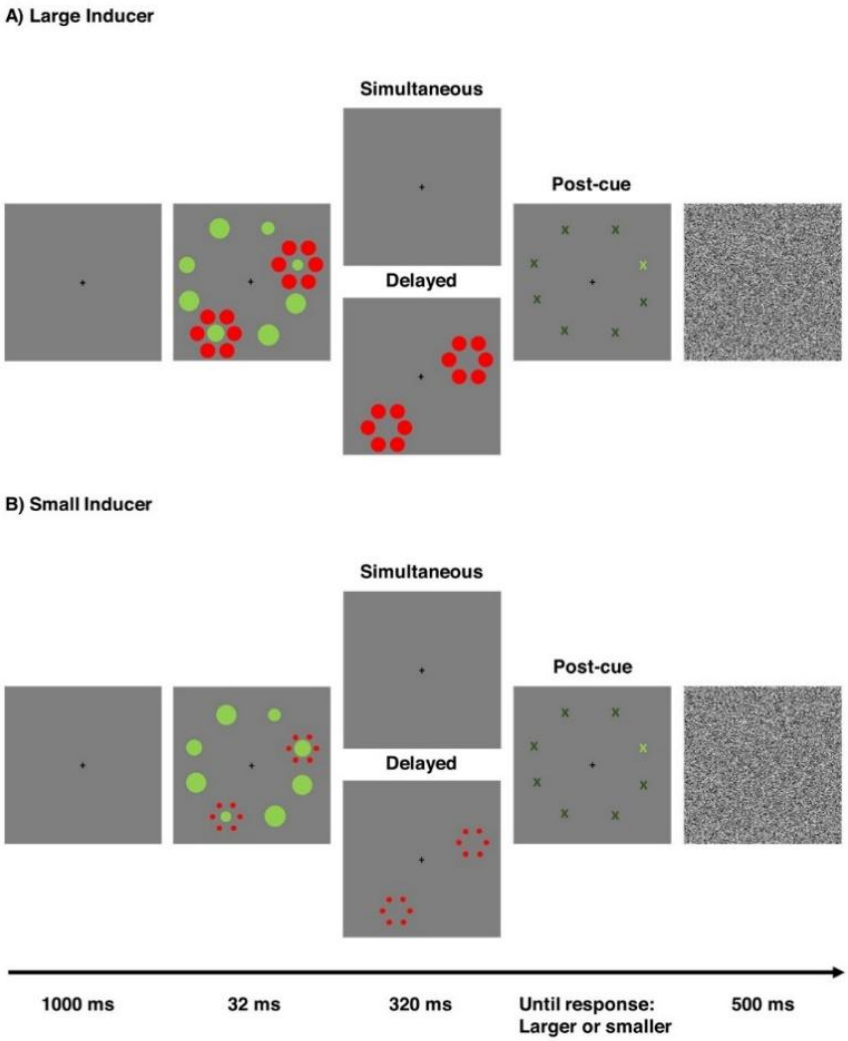

Fig. 3. Illustration of the task in Experiment 2. *Participants were asked to indicate whether the target circle, marked by a post-cue, was large or small. Two out of eight circles were surrounded by either large (A) or small (B) inducers. The post-cue indicated which of the two circles was the target. In the simultaneous viewing condition, all stimuli disappeared at the*

same time. In contrast, in the delayed viewing condition, the mask remained visible for an additional 320 ms after the test circles had disappeared.

Procedure

We manipulated the viewing type (simultaneous, delayed) to test the efficiency of OSM, using a paradigm similar to that of Choo and Franconeri (2010). Participants were presented with a size-discrimination task where they were supposed to identify the size of the target. Given that there were only two size categories, we asked them to report whether the target indicated by a post-cue was a large or a small target. Each trial started with a 1000 ms fixation period (Fig. 3). Following this, the distractors, target, non-target, and inducers were presented for 32 ms. Afterward, either the fixation cross alone (Fig. 3A-B, Simultaneous) or both the fixation cross and inducers (Fig. 3A-B, Delayed) remained on the screen for 320 ms. A post-cue then appeared, staying visible until participants pressed a button to indicate whether the target circle marked by the post-cue was large or small. At the end of each trial, a grey noise pattern was displayed for 500 ms. Each participant completed 640 trials (2 inducer types × 2 viewing conditions × 2 target sizes × 2 nontarget sizes × 40 repetitions). This experiment took around 40 minutes.

Statistical analyses

The percentage of correct responses was used as an outcome variable in Experiment 2. We performed a 2 x 2 x 2 x 2 ANOVA to test the effectiveness of OSM in our experimental paradigm. This analysis involved examining the impact of viewing condition (simultaneous, delayed), inducer type (small, large), target size (small, large) and congruency (congruent, incongruent). While the latter was irrelevant to our hypothesis, it was added to explain variance induced by target congruency.

Results

Figure 4 shows the percentages of correct responses with standard errors for within-subject contrast for the simultaneous and delayed viewing conditions under different inducer-type conditions. An ANOVA revealed the main effects of viewing condition [simultaneous vs. delayed] (F (1, 28) = 117.46, $p < .001$, $\eta^2_p = 0.808$), inducer type [small vs. large] (F (1, 28) = 9.76, $p = .004$, $\eta^2_p = 0.258$) and congruency [congruent vs. incongruent] (F (1, 28) = 10.86, $p = .003$, $\eta^2_p = 0.280$). However, the main effect of target size did not reach significance ($p = .313$). Additionally, significant interactions were observed between target size and inducer type (F (1, 28) = 17.40, $p < .001$, $\eta^2_p = 0.383$), as well as between inducer type and congruency (F (1, 28) = 7.18, $p = .012$, $\eta^2_p = 0.204$).

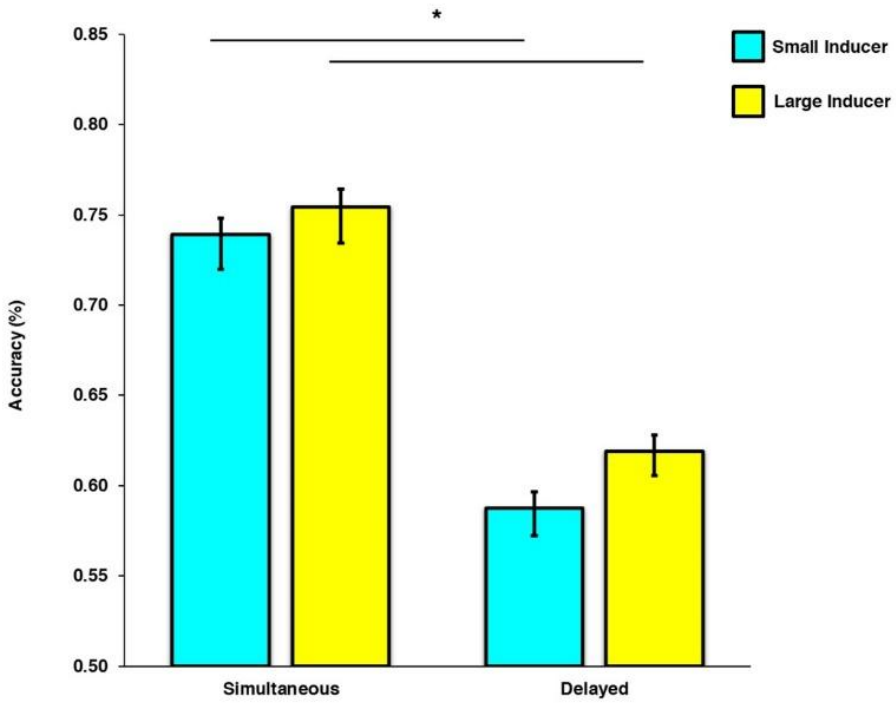

Fig. 4. The percentages of correct responses in Experiment 2. *The percentage of correct responses was plotted against inducer type and viewing condition. Blue bars represent the small inducers, and yellow bars represent the large inducers. Asterisks (*) represent significant differences at p < .05. Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014).*

A robust masking effect was revealed by a sharp decrease in the percentage of correct responses in the delayed viewing condition (*M* = 0.60, *SE* = 0.02) compared to those in the simultaneous viewing condition (*M* = 0.75, *SE* = 0.02). Interestingly, performance demonstrated superiority in the conditions with large inducers (*M* = 0.69, *SE* = 0.02) over the conditions with small inducers (*M* = 0.66, *SE* = 0.02), indicating a stronger masking effect induced by the small inducer. Additionally, the interaction between target size and inducer type revealed that the impact of inducer type was present only when the target size was small. Specifically, small targets in the large inducer conditions (M = 0.70, SE = 0.03) were detected more accurately than in the small inducer conditions (M = 0.62, SE = 0.03). However, such an effect was not observed when the target size was large.

Incongruent trials (*M* = 0.70, *SE* = 0.02), where the target and nontarget had different sizes, generated more accurate responses than congruent trials (*M* = 0.65, *SE* = 0.02). However, inducer type and congruency interaction revealed that the size of the inducer produced a difference only for the congruent trials. Specifically, the percentage of correct responses was greater in the large inducer conditions (M = 0.67, SE = 0.03) than in the small inducer conditions (M = 0.63, SE = 0.03) only if the target and nontarget were the same size. Overall, these findings indicate that OSM efficiently reduced the reportability of the masked circle size.

2.2.5. Experiment 3: Size averaging

Experiment 3 tested whether the size averaging task can effectively capture the illusory size changes induced by the Ebbinghaus illusion. Employing two distinct target sizes (small and large) that reflected the induced perceived size changes observed in Experiment 1, we explored whether our size averaging task can detect such alterations in target size.

Method

Participants

The same group of twenty-nine participants (M = 29.38 years, SD = 5.52, 17 females) attended Experiment 3.

Stimuli

Eight heterogeneously sized green circles were presented around the fixation cross. Unbeknown to the participants, three green circles were assigned as the target stimuli (e.g., Fig. 6). For each participant, these three targets were presented with a size matching the PSE values estimated in Experiment 1. Namely, all three targets were either presented at the stimulus's perceived size surrounded by large inducers (small target size) or presented at the stimulus's perceived size surrounded by small inducers (large target size) in Experiment 1. The average size of the five green distractor circles was always 0.9 degrees of visual angle, and the average size was calculated based on the normal cumulative distribution (M = 0.90, SD = 0.15). Constant stimuli were used to detect the perceived average size of the green circles. The comparison circle's size varied around the average screen size with .1 degrees of visual angle increments, resulting in two different comparison size lists. There were 10 different comparison sizes and each comparison circle was presented 10 times.

Procedure

Each trial started with a 500 ms fixation (Fig. 5). Then, eight green circles were presented for 32 ms, followed by a 320 ms fixation. In the response window, a comparison circle was displayed until participants pressed a button to indicate whether the comparison circle was larger or smaller than the average size of all eight circles. Each trial ended with a grey noise pattern for 500 ms. Participants performed 200 trials (2 target sizes × 10 comparison sizes × 10 repetitions). This experiment took around 15 min.

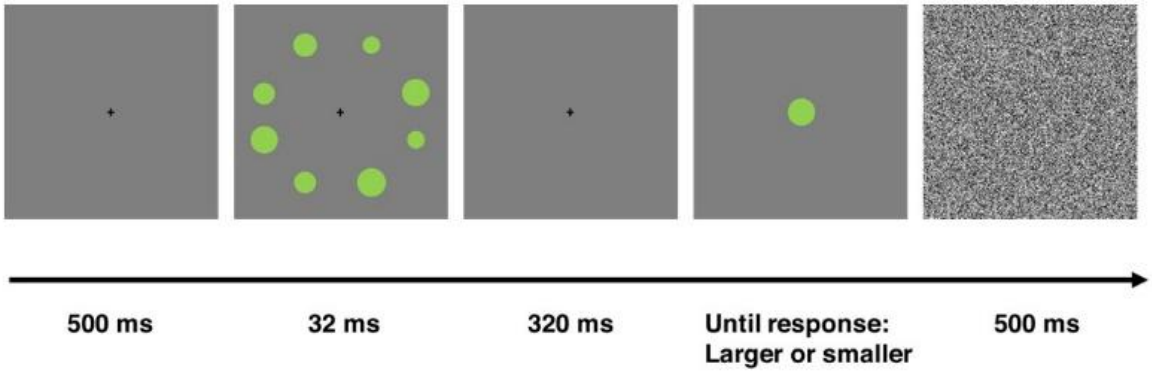

Fig. 5. Illustration of the task in Experiment 3. *Participants were asked to indicate whether the comparison circle (time window 4) was larger or smaller than the average size of all eight circles (time window 2).*

Statistical analyses

To assess how participants perceived the average size of the stimuli, we generated psychometric curves for each participant across experimental conditions. The PSE represented the point where the comparison circle was perceived at an equivalent size to the stimuli's average size. Curve widths reflected the level of uncertainty in the participants' responses, with wider curves indicating greater perceptual uncertainty. We calculated goodness of fit measures when fitting psychometric curves to the data. The obtained curves demonstrated a strong fit in Experiment 3 (r ranged between .833 and .995). The perceived average size of all green circles was used as an outcome variable. Given our directional hypothesis, we conducted a one-tailed paired sample t-test to assess how changes in target size (small and large) affected the perceived average size.

Results

Figure 6 depicts the mean PSEs with standard errors for within-subject contrast for the small and large average sizes. The results indicated a significant difference in the mean PSE values

for small and large targets (t (28) = 1.94, p = .032, Cohen's d = 0.359), while no significant difference was found between the curve widths (t (28) = 1.41, p = .170, Cohen's d = 0.262). Specifically, participants estimated the average size as significantly smaller in the small target condition (M = 0.80, SE = 0.02) than in the large target condition (M = 0.83, SE = 0.03), showing that the size averaging paradigm was sensitive enough to detect size changes induced by the Ebbinghaus illusion (Experiment 1).

Essentially, these findings suggest that the presence of small targets (the mean PSE of the large inducer per participant detected in Experiment 1) could reduce the perceived size of the entire stimulus, while the presence of large targets (the mean PSE of the small inducer per participant detected in Experiment 1) could increase it.

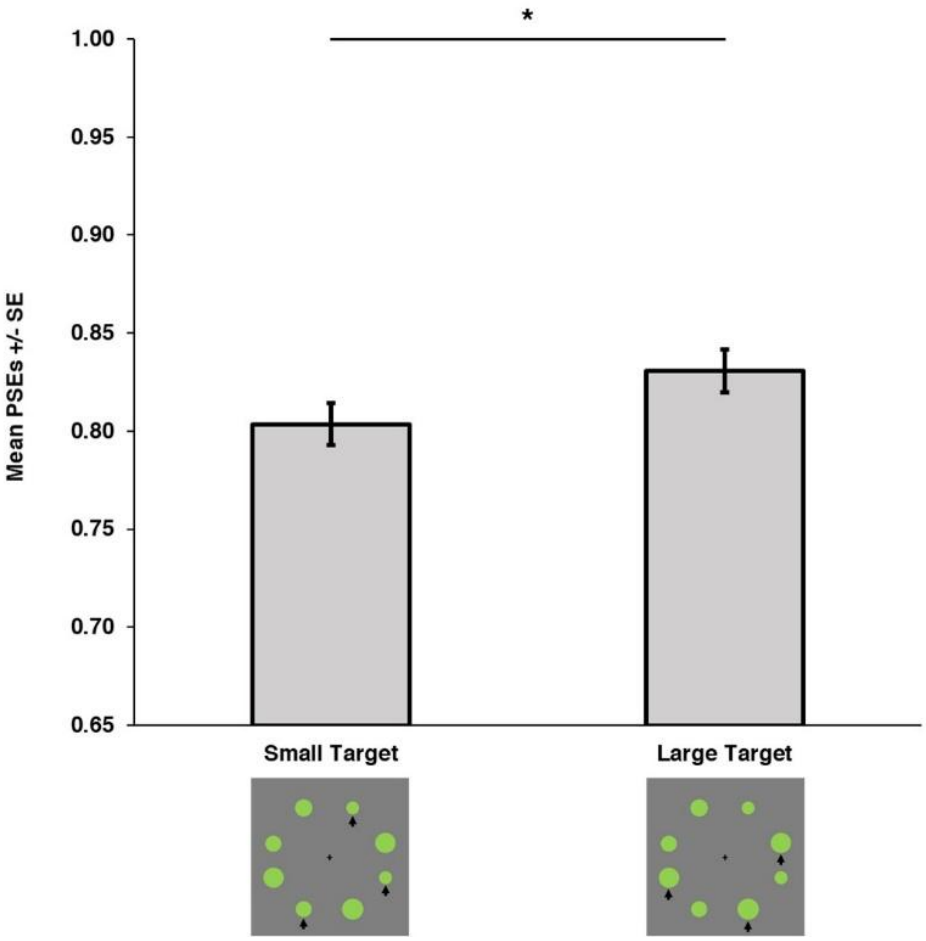

Fig. 6. Perceived average size across different target sizes in Experiment 3. *Averaged PSEs were plotted for small and large target size. Asterisks (*) represent significant differences at $p < .05$. Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014). Black arrows (↑) illustrate an example of randomly assigned three target positions across experimental conditions. Please note that these arrows were not part of the original display.*

2.2.6. Experiment 4A: Size rescaling and masking

The results of Experiments 1-3 indicated that our stimulus configurations were suitable to generate a robust Ebbinghaus illusion and an efficient OSM. Likewise, the size averaging paradigm was sensitive enough to reveal size changes with a magnitude as induced by our

variant of the Ebbinghaus illusion. In Experiment 4A, we combined these three paradigms to reveal the effect of context integration on hidden size averaging.

Method

Participants

One participant from the group of twenty-nine participants couldn't take part in Experiment 4A-B. So, the remaining twenty-eight participants (M = 29.32 years, SD = 5.62, 17 females) took part in Experiments 4A-B. We calculated goodness of fit measures when fitting psychometric curves to the data (see below in the Statistical Analysis Subsection). Participants with goodness of fit values smaller than $r = 0.63$ (corresponding to a $p > 0.05$) were removed from the sample. In the final sample, there were 23 participants (M = 30 years, SD = 5.83, 13 females).

Stimuli

Three out of eight circles were randomly chosen as target circles and surrounded by small or large inducers that allowed them to act as Ebbinghaus inducers and masking stimuli inducing OSM (Fig. 7A-B). All three targets were always presented at 0.9 degrees of visual angle. Accordingly, every change in the perceived average size could be attributed to an illusion effect induced by the Ebbinghaus illusion. The physical average size of the stimulus display (excluding inducers) was either 0.9 or 1.1 degrees of visual angle (e.g., Fig. 8, horizontal dashed grey lines). This was achieved by setting the average size of the five distractors to 0.9 or 1.2 degrees of visual angle. The average size of the five distractors was calculated based on the normal cumulative distribution, and the size of each circle was varied with a standard deviation of 0.15 degrees of visual angle around the mean (0.9 or 1.2 degrees of visual angle). The target size was kept constant, and only the distractor average size was manipulated. This way, we ensured that participants averaged all green circles on the screen without

strategically selecting a subset. Specifically, if participants performed the task correctly, then their reported perceived average size is supposed to be larger in conditions with a large distractor average than those with a small distractor average.

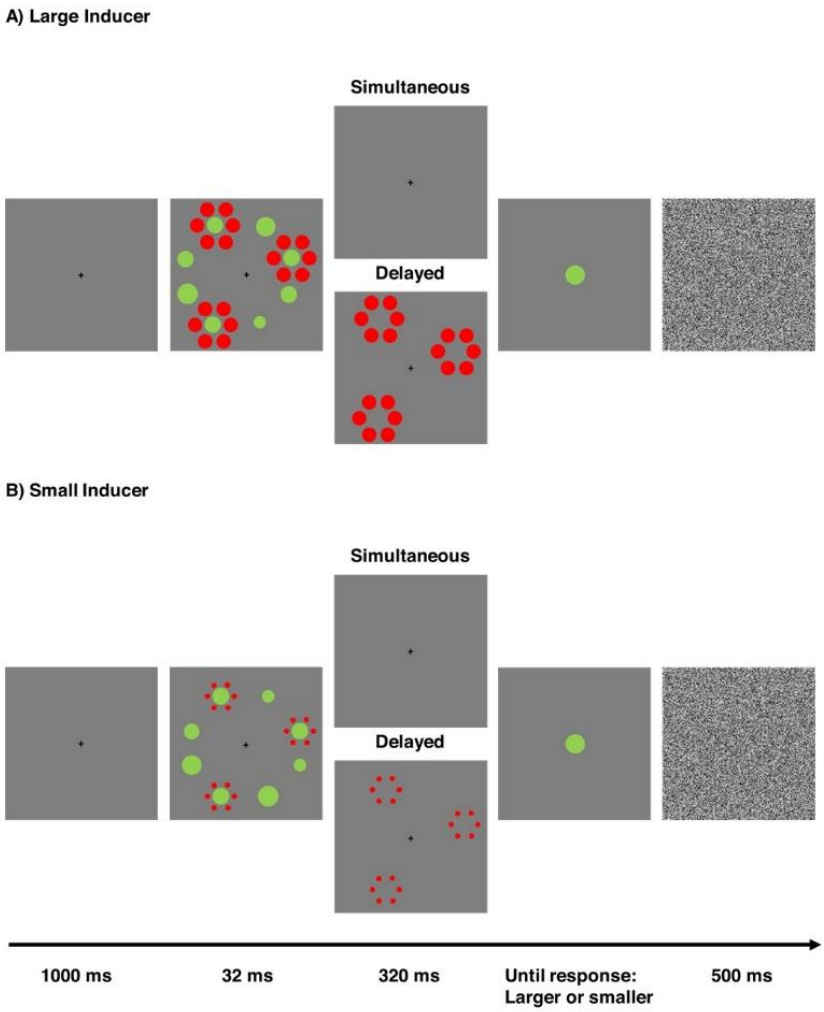

Fig. 7. Illustration of the task in Experiment 4A. *Participants were asked to indicate whether the comparison circle (time window 4) was larger or smaller than the average size of all green circles (time window 2). Three out of eight circles were surrounded by either large (A) or small (B) inducers. In the simultaneous viewing condition, all stimuli disappeared at the same time. In contrast, in the delayed viewing condition, the mask remained visible for an additional 320 ms after the test circles had disappeared.*

The targets surrounded by the three inducers were never adjacent and always separated by at least one distractor circle. Furthermore, targets were present on every trial in both hemifields, one in the left and two in the right hemifield or vice versa. Constant stimuli were used to detect the perceived average size of the green circles. The size of the comparison circle varied around the physical average size of the entire stimulus display with .1 degrees of visual angle increments. Accordingly, different comparison size lists were used for the two average size conditions. In each block, there were 10 different comparison sizes (half of them were smaller than the average size of all green circles, and the other half was larger), and each comparison circle was presented 10 times.

Procedure

The procedure of Experiment 4A was identical to Experiment 2, except for the following differences. First, in Experiment 4A, three targets were surrounded by inducers, compared to just one target in Experiment 2. Second, participants were asked to decide whether the average size of all green circles was smaller or larger than that of the comparison circle. Finally, Experiment 4A included not only simultaneous, but also delayed viewing condition.

Each trial started with a 1000 ms fixation, followed by a test screen consisting of green circles and inducers for 32 ms (Fig. 7). In the simultaneous viewing condition, all stimuli disappeared at the same time, and only the fixation cross appeared on the screen for 320 msec (Fig. 7A-B, Simultaneous). In the delayed viewing condition, the target circles disappeared, but the inducers persisted on the screen for 320 ms (Fig. 7A-B, Delayed). A comparison circle was presented at the centre of the screen until participants pressed a button to indicate whether the comparison circle was larger or smaller than the average size of all green circles. A grey noise pattern was presented for 500 ms to indicate the end of a trial. Participants completed 800 trials (2 inducer types × 2 viewing conditions × 2 distractor

average sizes × 10 comparison sizes × 10 repetitions). This experiment took around 50 minutes.

Eye-tracking data acquisition

We collected eye movement data using Eyelink 1000 (SR Research, Mississauga, Ontario, Canada) to verify that participants maintained their gaze on the fixation cross throughout Experiments 4A-B. Eye movement data were recorded from the right eye at a sampling rate of 500 Hz. Before the experiment, participants performed a five-point calibration and validation procedure. A three-times bigger circular area (1.2 degrees of visual angle) around the fixation cross was defined as a region of interest (ROI), and we then calculated the time participants spent within this ROI throughout the experiment. Eye movement data were recorded throughout the entire experiment, but only critical periods were further analysed (Fig. 7, time windows 1-2-3). The average coordinates of the fixation cross for each trial were employed as a drift check for that specific trial. Preprocessing of the eye movement data was performed using RStudio Version 4.2.0 (RStudio Team, 2015).

Statistical analyses

As in Experiment 3, we assessed how participants perceived the average size of the stimuli by fitting psychometric curves to each participant's data for each experimental condition. The PSE represented the point where the comparison circle was perceived at an equivalent size to the stimuli's average size. Curve widths reflected the level of uncertainty in the participants' responses for each experimental condition, with wider curves indicating greater perceptual uncertainty. We calculated goodness of fit measures when fitting psychometric curves to the data. The obtained curves demonstrated a strong fit in Experiment 4A (r ranged between .743 and .997). The perceived average size of all green circles was used as an outcome variable. A 2 x 2 x 2 ANOVA was conducted to detect the illusory effects on hidden size averaging, with

the factors viewing condition (simultaneous, delayed), inducer type (small, large) and distractor average size (small, large). The estimated curve widths were analysed in a separate 2 x 2 x 2 ANOVA with viewing condition (simultaneous, delayed), inducer type (small, large), and distractor average size (small, large) as within-subjects factors.

Results

Eye movement data

Due to technical problems, we could not use two participants' eye movement data. Following the exclusion of five outliers based on *r* values (see above in the Participants Subsection), the analysis of eye movement data involved the remaining twenty-one participants. A 2 x 2 x 2 ANOVA was employed to examine the percentage of time that participants maintained fixation on the ROI in each experimental condition, with the factors of inducer type (small, large), viewing condition (simultaneous, delayed) and distractor average size (small, large). During the experiment's critical periods, participants spent an average of 95.45% of their time within the ROI. ANOVA did not reveal significant main effects of viewing condition [simultaneous vs. delayed] (F (1, 20) = 0.77, p = .391, η^2_p = 0.037), inducer type [large vs. small] (F (1, 20) = 2.44, p = .134, η^2_p = 0.109) and distractor average size [small vs. large] (F (1, 20) = 0.44, p = .513, η^2_p = 0.022). Besides, none of the interactions reached significance (all $ps \geq .052$), indicating that participants maintained their gaze similarly across experimental conditions.

Analysis of the perceived average size

Figure 8 represents the mean PSEs with standard errors for within-subject contrast for the viewing condition and inducer type, in the (A) small and (B) large distractor averages. ANOVA revealed significant main effects of viewing condition [simultaneous vs. delayed] (F (1, 22) = 8.50, p = .008, η^2_p = 0.279), inducer type [large vs. small] (F (1, 22) = 51.92, p <

.001, η^2_p = 0.702) and distractor average size [small vs. large] (F (1, 22) = 49.70, p < .001, η^2_p = 0.693). Furthermore, an interaction between inducer type and distractor average was observed (F (1, 22) = 19.02, p < .001, η^2_p = 0.464).

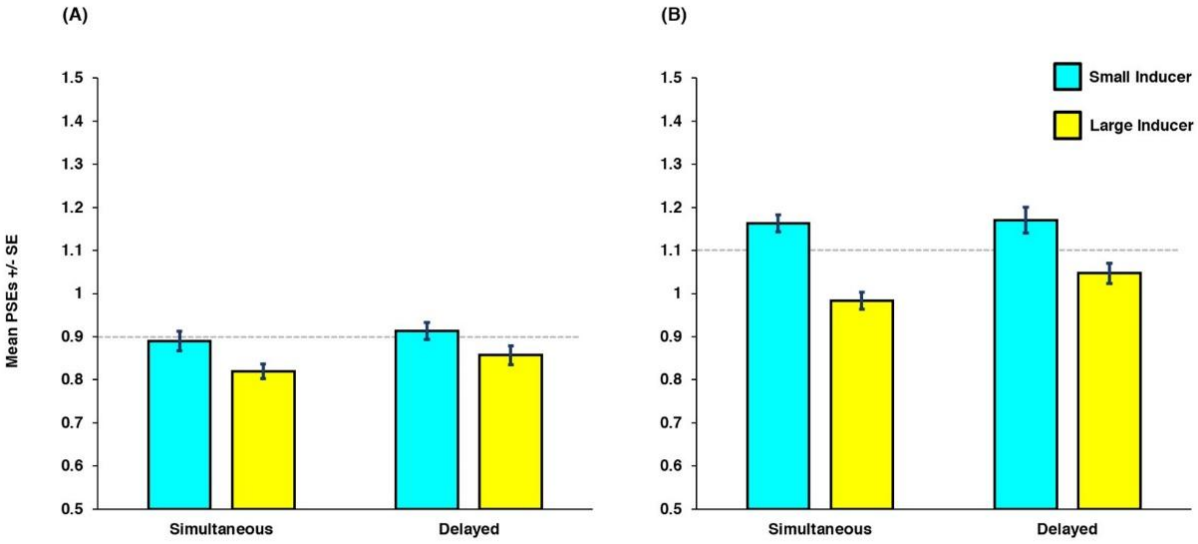

Fig. 8. Perceived average size for small (A) and large (B) distractor averages in Experiment 4A. *Averaged PSEs were plotted against the inducer type and viewing condition. Blue bars represent the small inducers, while yellow bars indicate the large inducers. Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014). The horizontal dashed grey lines represent the physical average size of the stimulus display.*

Most importantly, participants estimated the average size as larger in the small inducer conditions (M = 1.03, SE = 0.03) than in the large inducer conditions (M = 0.93, SE = 0.02), indicating the effect of contextual information. Given that no significant interaction was observed between inducer type and viewing condition (F (1, 22) = 1.50, p = .234, η^2_p = 0.064), the data suggest that the effect of contextual information did not differ between

masked and non-masked objects (Fig. 8A-B). Moreover, the mean PSEs were overestimated in the delayed viewing condition (M = 1.00, SE = 0.02) compared to the simultaneous viewing condition (M = 0.96, SE = 0.03) (Fig. 8A-B).

As expected, we observed a notable difference between the small and large distractor averages. Specifically, the averaged PSEs in the large distractor average condition (M = 1.09, SE = 0.03) were greater than the ones in the small distractor average condition (M = 0.87, SE = 0.03) (Fig. 8A-B).

Curve Widths

The estimated curve widths were analysed by conducting a 2 x 2 x 2 ANOVA with the factors viewing condition (simultaneous, delayed), inducer type (small, large) and distractor average size (small, large). ANOVA revealed neither main effects (all ps ≥ .188) nor interactions (all ps ≥ .077).

2.2.7. Experiment 4B: Size rescaling or size contrast?

The results from Experiment 4A could be attributed to a context-based modulation of the target items. However, one could argue that the inducers generated a more general effect on elements present in the display other than the target stimuli. The goal of Experiment 4B was to investigate whether the observed effect in Experiment 4A could be attributed to the size rescaling of masked or non-masked target objects or whether it signifies a more general influence of Ebbinghaus inducers. Specifically, we explored whether the apparent size of one set (e.g., Ebbinghaus inducers) would influence the perceived average size of the other set (e.g., distractor circles), indicating a size-contrast effect which may represent a broader impact of inducers on other elements on the screen. To explore this, we conducted a control experiment identical to Experiment 4A, with the only distinction being the absence of targets within the inducers. If the results from Experiment 4A were independent of target-related size-rescaling and instead stemmed from the more general effects of inducers, then we would

anticipate identical results in Experiment 4B. Specifically, the obtained results would not differ based on the presence or absence of targets. However, if the size rescaling induced by Ebbinghaus inducers altered the target size representation, we would expect to observe an additional modulation by the Ebbinghaus inducers in Experiment 4A compared to Experiment 4B.

Method

Participants

The same twenty-eight participants who attended Experiment 4A took part in Experiment 4B. We calculated goodness of fit measures when fitting psychometric curves to the data (see below in the Statistical Analysis Subsection). Participants with goodness of fit values smaller than $r = 0.63$ (corresponding to a $p > 0.05$) were removed from the sample. In the final sample, there were 23 participants (M = 30 years, SD = 5.83, 13 females).

Stimuli

All stimuli used in Experiment 4B were the same as in Experiment 4A. However, the only difference was the absence of target circles within the inducers. Hence, the average size of the overall screen was either 0.9 or 1.2 degrees of visual angle corresponding to the distractor average size since there were no more target circles to influence the overall screen average in the conditions with a large distractor average (e.g., Fig. 9, horizontal dashed grey lines).

Procedure

The procedure was identical to the one in Experiment 4A.

Eye-tracking data acquisition

We applied the same protocol to collect eye movement data.

Statistical analyses

The obtained curves demonstrated a strong fit in Experiment 4B (r ranged between .772 and .995). The perceived average size of all green circles was used as the dependent variable. ANOVA with the mean PSE values was performed with the factors viewing condition (simultaneous, delayed), inducer type (small, large) and distractor average size (small, large). After that, we applied a normalisation procedure to facilitate a comparable assessment of the PSE values across different distractor averages. This normalisation was necessary due to the differences in the physical average size of the green circles in the large distractor average condition between Experiment 4A and Experiment 4B. In Experiment 4A, we employed a constant target size (0.9 degrees of visual angle) and manipulated the distractor average size (0.9 and 1.2 degrees of visual angle), whereas no targets were presented in Experiment 4B. This resulted in different average sizes for the large distractor conditions. Consequently, in Experiment 4A, the physical average size of the green circles was 1.1 degrees of visual angle, calculated as follows: $((5 \times 1.2) + (3 \times 0.9)) / 8 = 1.1$. In contrast, the physical average size of the green circles was 1.2 degrees of visual angle in Experiment 4B, calculated as follows: $(5 \times 1.2) / 5 = 1.2$. To normalise the raw PSE values, we calculated each PSE value as a percentage relative to the particular physical average size. This procedure accounts for variations in the distractor's average size. For example, for a given PSE value (e.g., 1.1) and a screen average (e.g., 0.9), the normalised value was computed using the formula: (PSE × 100) / Screen Average. Then, we subtracted target-absent PSE percentages (Experiment 4B) from target-present PSE percentages (Experiment 4A). This subtraction enables us to observe the effect of the target’s presence on the PSE values.

Results

Eye movement data

This analysis included the same twenty-one participants as those in Experiment 4A. ANOVA was conducted to assess the percentage of time participants spent on ROI in different conditions, with the factors of inducer type (small, large), viewing condition (simultaneous, delayed) and distractor average size (small, large). On average, participants spent 95.79% of their time within the ROI. ANOVA did not show significant main effects of viewing condition [simultaneous vs. delayed] (F (1, 20) = 2.13, p = .160, η^2_p = 0.096), inducer type [large vs. small] (F (1, 20) = 2.07, p = .165, η^2_p = 0.094) and distractor average size [small vs. large] (F (1, 20) = 0.83, p = .373, η^2_p = 0.040). Similar to Experiment 4A, no significant interactions were observed (all $ps \geq .329$). Accordingly, fixation was comparable across the various experimental conditions.

Analysis of the perceived average size

Figure 9 represents the mean PSEs with standard errors for within-subject contrast for the viewing condition and inducer type in the small (A) and large (B) distractor averages. ANOVA revealed significant main effects of both inducer type [small vs. large] (F (1, 22) = 24.49, $p < .001$, η^2_p = 0.527) and distractor average size [small vs. large] (F (1, 22) = 60.00, $p < .001$, η^2_p = 0.732). However, the main effect of viewing condition did not reach significance. Moreover, an interaction between inducer type and distractor average was detected (F (1, 22) = 21.71, $p < .001$, η^2_p = 0.497).

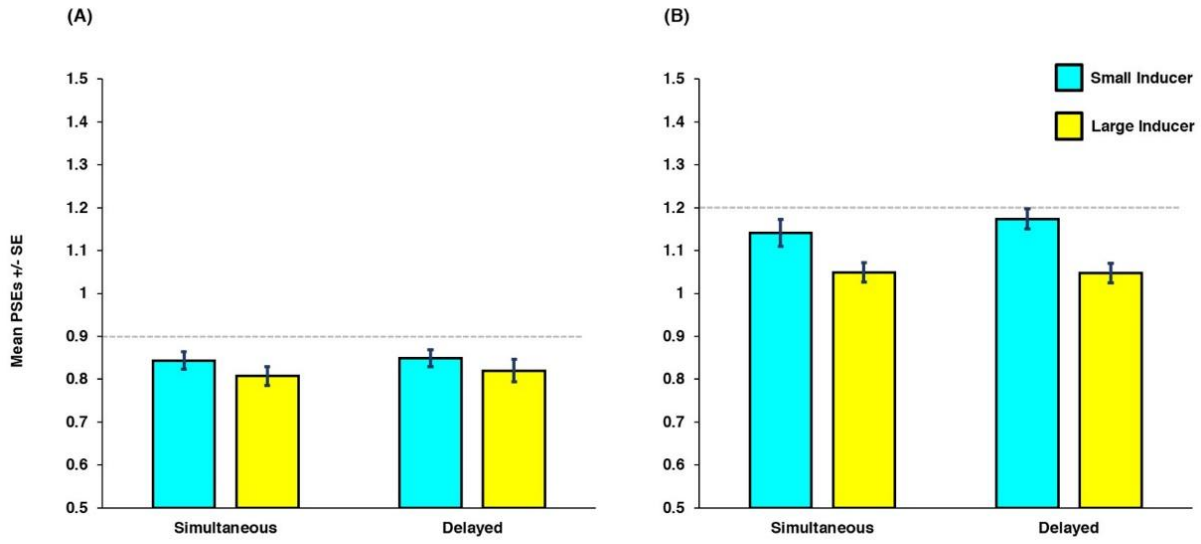

Fig. 9. Perceived average size for small (A) and large (B) distractor averages in Experiment 4B. *Averaged PSEs were plotted against the inducer type and viewing condition. Blue bars represent the small inducers, while yellow bars indicate the large inducers. Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014). The horizontal dashed grey lines represent the physical average size of the stimulus display.*

As in Experiment 4A, where the targets were present, the average size was estimated as larger in the small inducer conditions (M = 1.00, SE = 0.03) compared to the large inducer conditions (M = 0.93, SE = 0.03) (Fig. 9A-B). Also, the averaged PSEs in the large distractor average condition (M = 1.10, SE = 0.03) were greater than in the small distractor average condition (M = 0.83, SE = 0.03) (Fig. 9A-B). Additionally, the inducer type and distractor average interaction showed that the difference between the small and large inducers was significant for the large distractor average. This pattern is consistent with the notion of a more general effect of the Ebbinghaus inducers.

However, we explicitly compared the results of the two experiments in one analysis to test for additional modulations induced by Ebbinghaus inducers in Experiment 4A.

Exploring on top effects: The inducer effect on the target

Figure 10 represents the normalised PSEs with standard errors for within-subject contrast for the viewing condition and inducer type. Experiment 4B was conducted to estimate the general effects of the Ebbinghaus stimuli on the results of Experiment 4A. While both experiments involved the effects of inducers on distractor stimuli, only Experiment 4A involved additional effects of inducers on the masked and non-masked target items. Subtracting the target-absent PSE values (Experiment 4B) from the target-present PSE values (Experiment 4A), we could reveal target-related modulations. Therefore, we directly compared Experiments 4A and 4B. After normalising the raw PSE values by calculating the percentage of each PSE value considering the physical average size of the screen, we subtracted the target-absent PSE values (Experiment 4B) from the target-present PSE values (Experiment 4A). The resulting values were included in the analysis. A 2 x 2 x 2 ANOVA was conducted with the factors viewing condition (simultaneous, delayed), inducer type (small, large) and distractor average size (small, large). ANOVA revealed significant main effects of inducer type [small vs. large] (F (1, 22) = 6.52, p = .018, η^2_p = 0.229) and distractor average size [small vs. large] (F (1, 22) = 4.52, p = .045, η^2_p = 0.170). However, neither the main effect of viewing condition (p = .191) nor the interactions (all $ps \geq .082$) reached statistical significance.

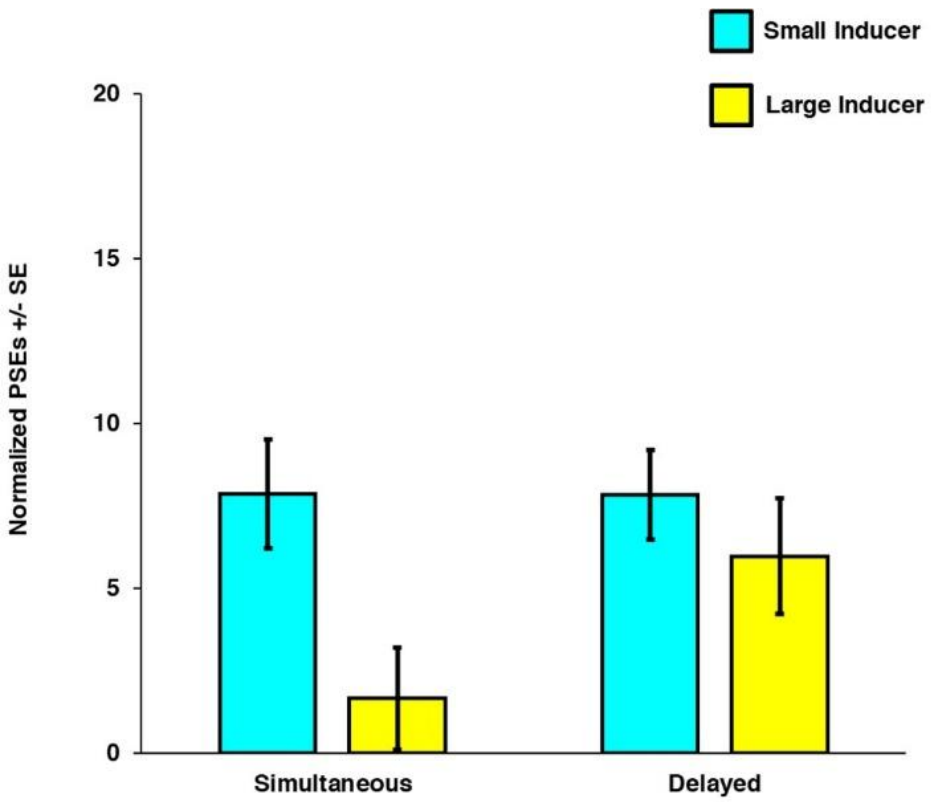

Fig. 10. Difference between Experiments 4A and 4B. *Normalised PSEs were plotted against inducer type and viewing condition. Blue bars represent the small inducers, while yellow bars indicate the large inducers. Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014).*

Most importantly, participants estimated the average size as larger in the small inducer conditions (M = 7.86, SE = 1.10) relative to that in the large inducer conditions (M = 3.82, SE = 1.28), indicating evidence of target-based size modulations (Fig. 10). Since, the interaction between the factors inducer type and viewing condition was not significant, we did not have evidence that this effect was different for masked and non-masked targets (F (1, 22) = 2.01, p = .170, η^2_p = 0.084). Furthermore, as expected, the average size was estimated larger in the large distractor average conditions (M = 7.26, SE = 1.27) than in the small distractor average conditions (M = 4.42, SE = 0.94).

Curve Widths

The estimated curve widths were analysed by conducting a 2 x 2 x 2 ANOVA with the factors viewing condition (simultaneous, delayed), inducer type (small, large) and distractor average size (small, large). As in Experiment 4A, ANOVA revealed neither main effects (all $ps \geq .184$) nor interactions (all $ps \geq .533$).

2.2.8. Discussion

Through behavioural experiments, we investigated how contextual information influences the perceived size of masked and non-masked objects. A combined paradigm allowed the induction of an illusion via the Ebbinghaus inducers. The Ebbinghaus inducers also efficiently functioned as a mask, inducing object substitution masking. Efficient masking decreased target recognisability and hence rendered target item encoding more implicit. Finally, size averaging allowed us to indirectly assess the size representations of the target stimuli so that it was possible to infer the effect of size rescaling of both masked and non-masked visual stimuli in the same way. The results of Experiment 4A showed that large Ebbinghaus inducers decreased the reported average size, whereas small inducers led to an increase. This effect was independent of the target items being masked.

Moreover, in a control experiment (Experiment 4B), we tested for a more general inducer effect on the display configuration by removing all target items while keeping all other display features identical to Experiment 4A. Experiment 4B indeed showed a general effect of inducers on distractor items. We explicitly tested for specific target effects by directly comparing the results of Experiments 4A and 4B. This analysis showed that over and above the general effects found in Experiment 4B, there was a target-related modulation of size by Ebbinghaus inducers, which was not significantly different for masked and non-masked objects.

The interpretation of Experiments 4A's and B's findings critically depends on the effectiveness of each experimental manipulation combined in our paradigm. Therefore, in addition to the main experiment, we verified the effectiveness of each paradigm by conducting Experiments 1-3.

Experiment 1 demonstrated the efficiency of our experimental setup in altering the perceived size of objects. The results showed a substantial perceived size difference between small and large inducers compared to the control condition without inducers, indicating that our variant of the Ebbinghaus illusion generated a robust size illusion effect in Experiment 1. However, the no-inducer condition differs from the two illusion conditions in terms of the number of stimuli, which could explain why no significant increase in perceived average size was detected in the small inducer condition compared to the no-inducer condition. A variant with inducer stimuli of the same size as the target might have provided a more controlled baseline for comparison. Please note, that the no-inducer condition was not used in the subsequent experiments.

Furthermore, it was crucial to ensure that our masking procedure efficiently altered target recognisability to infer the effects of size-rescaling of masked vs. non-masked stimuli. We used OSM following the experimental setup previously introduced by Choo and Franconeri (2010). The results from Experiment 2 revealed that our stimulus configuration reliably induced OSM, as evident from a robust performance decline in the delayed viewing condition. In principle, the simultaneous presentation of targets and inducers may have decreased target recognisability, hence generating a masking effect itself. However, results from Experiment 2 show that accuracy was significantly reduced in the delayed viewing condition as compared to the simultaneous viewing condition hence indicating that our experimental set-up successfully altered target recognisability. We used the stimulus parameters proven efficient in Experiment 2 to ensure robust masking in Experiment 4A.

Lastly, we tested whether the size averaging paradigm was sensitive enough to detect perceived size changes in the magnitude of the ones induced by the Ebbinghaus illusion in Experiment 3. We employed the individual PSE values from Experiment 1 as small and large target sizes to achieve this. Specifically, we adjusted the physical sizes of the target items to align with the perceived sizes altered by the Ebbinghaus illusion in Experiment 1. We employed three target circles in a single display to enhance the impact of target size. Experiment 3 confirmed that both small and large targets significantly influenced the perceived display's average size, leading to a decrease and an increase in the perceived average size, respectively.

After successfully confirming the efficiency of each paradigm in Experiments 1-3, the objective of Experiment 4A was to integrate these three paradigms to reveal the effect of context integration on hidden size averaging. The findings of Experiment 4A demonstrated a modulation in the perceived average size, regardless of the object being masked or not. No notable distinctions were evident between the simultaneous (non-masked) and delayed (masked) viewing conditions. These findings replicate those reported by Choo and Franconeri (2010), emphasising that even when OSM highly impairs an object's visibility, its contribution to the average size remains intact and is comparable to that of an unmasked object. The results of Choo and Franconeri (2010) have been criticised in a study by Jacoby and colleagues (2013). The main criticism raised in this study was directed towards the mask positions, particularly the possibility of adjacent masks, which could potentially introduce an attentional effect. Besides, Choo and Franconeri (2010) found that the average influence of the two target circles was weaker in masked conditions compared to non-masked conditions, showing a bias toward smaller responses. Following the critique by Jacoby and colleagues (2013), this result could be interpreted as evidence that OSM disrupts size averaging. We addressed these concerns by implementing three masks to ensure a diffuse distribution of

attention mandatory for OSM (Enns & Di Lollo, 1997). Further, we included two distinct distractor averages in Experiment 4A, which allowed testing participants' task performance. Contrary to Jacoby et al.'s findings (2013), which demonstrated that OSM disrupts single object representation and impairs averaging performance, the findings of Experiment 4A strongly support the notion that even masked objects contribute to size averaging. The results of Experiment 4A align with a prior study showing that ensemble summary statistics vary depending on contextual information, and the perceived average range is formed after size-rescaling mechanisms (Markov & Tiurina, 2021). Notably, our results take these findings one step further, showing that even the size of masked objects is first rescaled and then incorporated into the averaged size. Interestingly, even though size information was implicitly coded and not explicitly reportable in the delayed viewing conditions, a bottom-up signal still revealed its influence on perceptual decisions. This is consistent with a previous study showing that neural activity patterns in occluded areas of early visual cortex provide key insights into the categorization and specific details of nearby images (Morgan et al., 2019). Despite V1 and V2 receiving only feedback from higher cortical areas and lacking a direct feedforward signal, information can still be extracted from these occluded areas. This suggests that category information can be effectively anticipated by the internal models in the early visual cortex.

Previously, Im and Chong (2009) combined size averaging and the Ebbinghaus illusion to test whether the reported average size reflects a stimulus's perceived or physical size. In contrast to our study, their stimuli were clearly visible and unmasked. Small and large Ebbinghaus inducers induced, similar to our study, contextual information. Following a reasoning similar to ours, the authors argued that if the perceived size of target objects contributes to the reported average size, there would be either an overestimation or an underestimation of the average size. Instead, if the physical size contributes to the reported

average size, no difference would emerge between small and large inducer blocks. Their results showed that the perceived size of the objects contributed to the reported average size. Thus, their data align with our findings, indicating that each target size is rescaled through contextual information and then contributes to the average size, consistent with Tiurina and Utochkin (2019). These findings collectively demonstrate that contextual information influences the objects before they contribute to ensemble summary statistics.

In light of the current findings, it is reasonable to argue that size averaging comes into play just after the size representations are rescaled through contextual information. This hypothesis is supported by the finding that conscious perception of the surrounding inducers or target item is not essential for the Ebbinghaus illusion, indicating a subconscious contextual modulation (Chen et al., 2018). In Chen et al.'s study (2018), the target items and the surrounding inducers were rendered invisible by continuous flash suppression and backward masking. Significant illusory size judgments were observed in both masking procedures, demonstrating that the illusory effect induced by the Ebbinghaus illusion persisted even when the target or surrounding inducers were invisible. In contrast, similar findings could not be detected for the Ponzo illusion, which relies on feedback mechanisms from higher-level visual areas (Schwarzkopf et al., 2011). Chen et al.'s findings indicate that subconscious contextual modulations already occur in early visual processing, consistent with the results from a functional imaging study indicating that neural activity in V1 represents the perceived size of the objects (Sperandio et al., 2012).

Even though the Ebbinghaus illusion is widely recognized for its association with activations in early visual processing regions, particularly V1, recent studies suggest that this illusion may involve brain regions beyond V1 (Wu et al., 2023; Chen et al., 2024). Wu and colleagues (2023) reported an important role of the parietal cortex for the Ebbinghaus illusion. Namely, the connection from the right V1 to the superior parietal lobule (SPL) was

found to predict the strength of the Ebbinghaus illusion. A stronger Ebbinghaus illusion after inhibitory rTMS to the right SPL might be explained by altered inhibitory feedforward connectivities from V1 to SPL. This effect aligns with the observed negative correlation between the strength of this connection and the magnitude of the Ebbinghaus illusion. Furthermore, Chen et al. (2024) demonstrated that the Ebbinghaus illusion is influenced by feedback projections from higher to lower visual areas, highlighting the role of top-down signals in modulating this illusion. The influence of feedback mechanisms on size rescaling was further demonstrated by Zeng and colleagues (2020). Their study, which applied TMS stimulation, revealed that there is a top-down influence on context integration, indicating that feedback mechanism from LOC modulates the altered perceived size representations in the early visual cortex.

Despite object recognisability being highly impaired by OSM in Experiment 4A, the perceived average size was larger in the delayed viewing than the simultaneous viewing condition. A similar tendency to overestimate the average size in the delayed viewing condition was observed in previous studies (Choo & Franconeri, 2010; Jacoby et al., 2013). Nevertheless, this effect was not evident when target items were removed in Experiment 4B. Two perspectives may account for this overestimation in the delayed condition compared to the simultaneous condition in Experiment 4A. Initially, visual information processing differs between the simultaneous and delayed viewing conditions in Experiment 4A. In the simultaneous condition, where targets are presented within the inducers, the perceptual process benefits from faster processing due to the immediate availability of visual signals. However, in the delayed condition, the absence of visual signals due to OSM necessitates a more prolonged accumulation of evidence, resulting in extended processing time. This prolonged processing time aligns with previous studies demonstrating that stimuli presented

or processed longer tend to be perceived as larger (Thomas & Cantor, 1976; Rammsayer & Verner, 2014).

Secondly, neuronal activity initially progresses from lower-level visual areas to higher-order visual areas in visual processing, referred to as feedforward processing (Krasich et al., 2022; Boehler et al., 2008). Once this initial feedforward activity reaches a specific visual area, the activation is sent back to lower-level visual areas through feedback connections (recurrent processing). These feedback connections are believed to convey predictions about incoming information, while any error in these predictions is transferred via feedforward connections, aiding to refine subsequent processing. OSM is thought to interfere with recurrent processing while preserving the initial feedforward signal (Harris et al., 2013; Boehler et al., 2008; Di Lollo et al., 2000; Enns & Di Lollo, 1997). Specifically, the longer presentation of mask in the delayed viewing condition results in additional feedforward processing of the mask-alone representation. This additional feedforward processing of the prolonged mask-alone presentation disrupts the recurrent processing of the stimulus-mask representation. Consequently, context-induced perceptual biases may have a stronger impact and, may enhance the impact of the Ebbinghaus illusion in the delayed viewing condition. Integrating these perspectives, we suggest that the observed overestimation in the delayed viewing condition in Experiment 4A arises from the prolonged processing time required for accumulating evidence and the interruption of recurrent processes. These explanations could also clarify why we did not observe overestimation in Experiment 4B, where the absence of targets removed potential processing differences between the simultaneous and delayed viewing conditions.

Interestingly, we observed a modulation of the perceived average size in the control experiment (Experiment 4B). This finding cannot be attributed to rescaling the target size, as demonstrated in Experiment 4A, nor to the influence of inducers contributing to average size.

If Ebbinghaus inducers were incorporated into the items to be averaged, an opposing effect would be anticipated. Specifically, the presence of large inducers would lead to an increase in the reported average, while small inducers would result in a decrease in the reported average. However, we observed the exact opposite pattern. The data suggest that the inducers directly act on the perceived average size in more general terms. The fact that large inducers decreased while small inducers increased perceived average size suggests a size contrast effect, which may arise on multiple levels of processing. It is possible that the size of the items to be judged is already biased at the initial perceptual level, where all stimuli are presented simultaneously. This interpretation is supported by psychophysical (Song et al, 2011; Nakashima & Sugita, 2018) and neuroimaging (Schwarzkopf & Rees, 2013; Schwarzkopf et al, 2011) studies suggesting that the Ebbinghaus illusion occurs on early levels within the visual system. Alternatively, the decision might be influenced at a subsequent stage, where the contrast effect becomes more pronounced, leading participants to judge the distractors relative to the inducers in Experiment 4B. This latter view is consistent with early accounts of the Ebbinghaus illusion such as the one suggested by Massaro & Anderson (1971). Based on this account, the bias observed in Experiment 4B might reflect a size-contrast effect occurring at a post-perceptual stage. In particular, the inducers may have served as standards or references when the distractor circles are judged (Massaro & Anderson, 1971).

The question then arises, whether the effects observed in Experiment 4A can be fully explained by the same mechanism, namely by a contrast effect of the inducers on the visible items on the screen. We addressed this question by comparing the data from Experiments 4A and 4B. We subtracted the size average judgments in the control Experiment 4B with its respective judgments in Experiment 4A to eliminate any size contrast effect by the inducers on non-target stimuli. The rationale behind this was that if the changes in size-averaging were

solely due to a general size-contrast effect, then the two experiments should show no differences. On the other hand, any observed difference could be attributed to a rescaling effect of the target. What we found were on-top effects in Experiment 4A over Experiment 4B, indicating target-related effects. In particular, we found a significant impact of inducer size, with larger reported size averaging values for small inducers and smaller size averaging values for large inducers, indicating target-size-rescaling. Importantly, we did not find any difference between viewing conditions. Hence, target-size-rescaling was not found to be different between explicitly and implicitly coded objects.

Two key considerations could be raised when interpreting our findings. Initially, one might question whether our results have been influenced by post-perceptual decision biases (Firestone & Scholl, 2016). While such concerns are typical in psychophysics, we implemented several precautions to minimise the impact of these biases. Participants were instructed to report their responses without using mental strategies and were unaware of the experimental manipulations (inducer types, viewing conditions, and distractor sizes), making it unlikely that differences between conditions were due to general decision biases. The randomisation of block order further reduced potential biases from prior knowledge, and participants' lack of familiarity with psychophysical methods (e.g., the Method of Constant Stimuli) decreased the likelihood of systematic bias. Finally, we used two distractor averages to evaluate task performance. The target size was consistently set at 0.9 degrees of visual angle, while the distractor average sizes were manipulated to be either small (0.9) or large (1.2) degrees of visual angle. This manipulation altered the overall average size of objects on the screen. Our findings demonstrated a clear upward shift in perceived average size when the distractor average size increased, which would not be expected to observe when the effects were purely due to decisional biases.

Subsequently, as our paradigm involves several circles, it could potentially induce crowding effects. However, the stimuli that are located close enough to generate crowding effects are separated by colour. It is known, that perceptual grouping is one of the key factors for crowding. Stimuli from different sets of colours are unlikely to generate crowding as demonstrated by Kennedy and Whitaker (2010). In their study, they manipulated chromaticity channels (red-green, blue-yellow) and demonstrated that crowding occurs only when the target and flankers shared the same colour. Furthermore, most crowding paradigms typically involve longer stimulus durations (≥150 ms) that allow a proper allocation of attention (Levi et al., 2002; Kooi et al., 1994; Kennedy & Whitaker, 2010; Freeman & Pelli, 2007; Greenwood & Parsons, 2020; Li et al., 2020), whereas our stimuli were presented for a notably brief duration of 32ms, making it unlikely for a crowding mechanism to occur. Finally, Parkes and colleagues (2001) demonstrated that crowding did not prevent participants from obtaining summary statistics, such as the average orientation of a cluster of items.

2.2.9. Conclusion

In summary, we identified two critical aspects related to size-rescaling in the context of size-averaging. First, we found that the Ebbinghaus inducers play a pivotal role in altering the perceived average size of distractor stimuli. When large inducers were present, the perceived average size decreased, while smaller ones had the opposite effect. The results indicate that the average size of different groups of objects are not independent, but mutually alter their perceived size in a contrast-like fashion. These findings align with earlier theories regarding the Ebbinghaus illusion (Massaro & Anderson, 1971), suggesting that the context circles within the illusion serve as standards for size judgment.

Second, our results revealed that the inducers consistently influenced their size representation even when stimuli were masked. This observation indicates that the size-

rescaling effects operate at the early stages of visual processing, preceding explicit stimulus encoding.

2.2.10. References

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological science*, *19*(4), 392-398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological science*, *12*(2), 157-162.

Boehler, C. N., Schoenfeld, M. A., Heinze, H. J., & Hopf, J. M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. Proceedings of the National Academy of Sciences, 105(25), 8742-8747.

Chen, L., Qiao, C., Wang, Y., & Jiang, Y. (2018). Subconscious processing reveals dissociable contextual modulations of visual size perception. *Cognition*, *180*, 259-267.

Chen, L., Wu, B., Yu, H., & Sperandio, I. (2024). Network dynamics underlying alterations in apparent object size. Brain Communications, 6(1), fcae006.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision research*, *45*(7), 891-900.

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. *Attention, Perception, & Psychophysics*, *72*(1), 86-99.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*(4), 481.

Ebbinghaus, H. (1902). Grundzuge der Psychologie (Vol. 1., Band 2. Theil.). *Leipzig: Viet & Co*.

Enns, J. T. & Di Lollo, V. (1997). Object Substitution: A new form of masking in unattended visual locations. *Psychological Science, 8*, 135-139.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers*, *28*(1), 1-11.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. Behavioral and brain sciences, 39, e229.

Freeman, J., & Pelli, D. G. (2007). An escape from crowding. Journal of vision, 7(2), 22-22.

Greenwood, J. A., & Parsons, M. J. (2020). Dissociable effects of visual crowding on the perception of color and motion. Proceedings of the National Academy of Sciences, 117(14), 8196-8202.

Harris, J. A., Ku, S., & Woldorff, M. G. (2013). Neural processing stages during object-substitution masking and their relationship to perceptual awareness. Neuropsychologia, 51(10), 1907-1917.

Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception, & Psychophysics*, *71*(2), 375-384.

Jacoby, O., Kamke, M. R., & Mattingley, J. B. (2013). Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 233.

Kennedy, G. J., & Whitaker, D. (2010). The chromatic selectivity of visual crowding. Journal of Vision, 10(6), 15-15.

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. Spatial vision, 8(2), 255-280.

Krasich, K., Simmons, C., O'Neill, K., Giattino, C. M., De Brigard, F., Sinnott-Armstrong, W., ... & Woldorff, M. G. (2022). Prestimulus oscillatory brain activity interacts with evoked recurrent processing to facilitate conscious visual perception. Scientific Reports, 12(1), 22126.

Levi, D. M., Klein, S. A., & Hariharan, S. (2002). Suppressive and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking. Journal of Vision, 2(2), 2-2.

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. Scientific reports, 10(1), 904.

Markov, Y. A., & Tiurina, N. A. (2021). Size-distance rescaling in the ensemble representation of range: Study with binocular and monocular cues. *Acta Psychologica*, *213*, 103238.

Massaro, D. W., & Anderson, N. H. (1971). Judgmental model of the Ebbinghaus illusion. *Journal of experimental psychology*, *89*(1), 147.

Morgan, A. T., Petro, L. S., & Muckli, L. (2019). Scene representations conveyed by cortical feedback to early visual cortex can be described by line drawings. Journal of Neuroscience, 39(47), 9410-9423.

Nakashima, Y., & Sugita, Y. (2018). Size-contrast illusion induced by unconscious context. Journal of vision, 18(3), 16-16.

O'Brien, F., & Cousineau, D. (2014). Representing error bars in within-subject designs in typical software packages. *The quantitative methods for psychology*, *10*(1), 56-67.

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.

Rammsayer, T. H., & Verner, M. (2014). The effect of nontemporal stimulus size on perceived duration as assessed by the method of reproduction. Journal of Vision, 14(5), 17-17.

RStudio Team. (2015). RStudio: Integrated development for R.Boston, MA: RStudio, Inc.

Schwarzkopf, D. S., & Rees, G. (2013). Subjective size perception depends on central visual cortical magnification in human V1. PloS one, 8(3), e60550.

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. Nature Neuroscience,14(1), 28–30.

Song, C., Schwarzkopf, D. S., & Rees, G. (2011). Interocular induction of illusory size perception. BMC neuroscience, 12, 1-9.

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. Nature Neuroscience,15(4), 540–542.

Thomas, E. A., & Cantor, N. E. (1976). Simultaneous time and size perception. Perception & Psychophysics, 19(4), 353-360.

Tiurina, N. A., & Utochkin, I. S. (2019). Ensemble perception in depth: Correct size-distance rescaling of multiple objects before averaging. *Journal of Experimental Psychology: General*, *148*(4), 728.

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision research*, *32*(5), 931-941.

Wu, B., Feng, B., Han, X., Chen, L., & Luo, W. (2023). Intrinsic excitability of human right parietal cortex shapes the experienced visual size illusions. Cerebral Cortex, 33(10), 6345-6353.

Zeng, H., Fink, G. R., & Weidner, R. (2020). Visual size processing in early visual cortex follows lateral occipital cortex involvement. Journal of Neuroscience, 40(22), 4410-4417.

2.3. Study II

***Memis, E**., Fink, G. R., & Weidner, R. (Unpublished manuscript, included in this dissertation and made publicly available via KUPS). Cross-ensemble size contrast in summary statistics: Neural and behavioural evidence.*

2.3.1. Introduction

The human visual system demonstrates remarkable parallel processing by rapidly forming ensemble representations, efficiently summarising or averaging object features across groups - such as size (Ariely, 2001), distance (Tiurina & Utochkin, 2019), and orientation (Parkes et al., 2001). For example, size averaging enables observers to rapidly (in less than 50 ms) and accurately estimate the average size of a set of stimuli, even when they are unable to report information about individual items (Ariely, 2001; Haberman & Whitney, 2007; Alvarez & Oliva, 2008; Allik et al., 2014). Notably, observers can accurately report the average size of a group of objects, even when the visibility of some objects is impaired by object-substitution masking (OSM) (Choo & Franconeri, 2010).

Size averaging performance remains resilient to changes in object features such as set size and set density (Chong & Treisman, 2005) and summary statistics are calculated simultaneously for multiple groups of objects without any drop in performance compared to sequential presentation. Overall, these findings indicate that the respective computations occur in parallel and at early stages of the visual processing hierarchy (Chong & Treisman, 2003).

Evidence from EEG suggests that ensemble representations such as average size are formed even before individual objects are processed (Epstein & Emmanouil, 2021). Although these findings argue that summary statistics are generated early on, the fact that summary statistics are calculated for different groups of objects in parallel (Chong & Treisman, 2005) implies that preattentive grouping processes have to occur first.

Taken together, these findings support the idea that individual objects and ensemble representations of objects are represented differently in the visual system and that information about ensemble representations for different object groups are available relatively early and simultaneously for different groups. It is yet unclear, whether such different summary statistics constitute independent descriptors or whether they are integrated at some level of processing where they might interact and whether such an interaction becomes perceptually relevant.

In a recent study, we found evidence that summary size statistics may not be coded independently but instead influence each other in a contrast-like fashion (Memis et al., 2025). In a size-averaging experiment we found that the average size of a set of objects was altered by the presence of stimuli inducing the Ebbinghaus illusion. Large Ebbinghaus inducers led to a decrease, whereas small inducers led to an increase in the perceived average size of the task-relevant set indicating instead a contrast-driven modulation. The pattern is consistent with the idea that one group of objects acts as the context for another set and that these contextual objects serve as benchmarks for size judgments, thereby producing a contrast-like effect (Massaro and Anderson, 1971).

In Memis et al.'s experiment, the Ebbinghaus inducers were all of the same size (either large or small) so that their average size and the size of a single inducer object were identical. This confound makes it difficult to uniquely determine whether the observed contrast effects were due to the overall summary size statistics or the specific sizes of the individual inducers. Therefore, in the present experiment, we used heterogeneous sets of stimuli with variable sizes, while keeping their average size constant to disentangle individual object size and group average size.

In particular, we employed behavioural measures and fMRI to investigate whether the perceived average size of one set of stimuli influences the perceived average size of another

set in a contrast-like fashion and whether such a contrast effect alters stimulus coding in early visual regions. Participants were presented with two sets of stimuli and were asked to report the perceived average size of a task-relevant set. Task-relevant and task-irrelevant sets were presented in different quadrants of the visual field, which allowed to separate their neural signatures in early retinotopic regions of the visual cortex. Furthermore, spatially separating the different groups of stimuli enabled us to quantify the effects of mutual size-contrasts on the early coding of objects. In addition, based on Murray et al.'s (2006) finding that increased perceived size correlates with larger representations in early visual regions, we hypothesised that: (1) neural activation in regions encoding task-relevant sets would be larger when surrounded by small task-irrelevant objects as compared to large ones; (2) neural activation in regions encoding task-irrelevant sets would be larger when surrounded by small task-relevant objects when compared to large ones; and (3) these neural patterns would correspond with behavioural results showing a size-contrast effect. Such mutual influence between the perceived average size of different sets would indicate that size-contrast effects, as observed in size-contrast illusions, may occur at the level of statistical summary descriptors.

2.3.2. Materials and methods

Participants

Twenty-nine healthy participants (M = 29.28 years, SD = 4.54, 11 females) took part in the fMRI experiment. All participants had normal or corrected-to-normal vision, and had normal colour vision as tested by the Velhagen-Broschmann pseudoisochromatic plates (Velhagen & Broschmann, 1997). Participants provided written informed consent prior to the experiment. They were remunerated for their time with 20 euros per hour for the fMRI session and 15 euros per hour for the behavioural session. The study was approved by the ethics committee of the German Society of Psychology (Application ID: WeidnerRalph2024-07-31-VA). The sample size was determined based on effect sizes observed in pilot studies to

make sure the current sample size would be sufficient to reveal significant differences between the experimental conditions (in a repeated-measures analysis of variance) with 95% power and an alpha level of 0.05. Considering the strong effect observed in the behavioural experiment, we cautiously increased the number of participants for the fMRI study to account for potential variations in the observed effect. Power estimates were computed using G*Power (Erdfelder et al., 1996).

Stimuli

The experiment was run via Visual Studio Code 1.68.1 using PsychoPy 2021.2.3 scripts. The experimental stimuli consisted of two uniformly coloured sets of circles, with eighteen green and eighteen red circles, presented on a grey background (49.90 cd/m²) at 40 degrees of visual angle over a black screen (Fig. 1). Participants' task was to indicate the perceived average size of a task-relevant set (e.g., green) while ignoring the task-irrelevant one (e.g. red). Due to sample size, full counterbalancing of the task-relevant colour was not feasible: 15 participants were assigned the green set as the task-relevant set, while 14 participants were assigned the red set. The luminance values were measured on the computer used to run the experiment, with green circles at 76.71 cd/m^2 and red circles at 18.27 cd/m^2.

In each trial, one quadrant of the display was randomly assigned as the target quadrant and included fourteen task-relevant circles (Fig. 1, time window 2). The quadrant diagonal to the target quadrant was assigned as the distractor quadrant, containing six task-irrelevant circles. The remaining two adjacent quadrants included six task-irrelevant and two additional task-relevant circles. This way, participants had to attend the whole screen to perform the task. The average size of the two additional circles in the adjacent quadrants was equal to the average size of the task-relevant circles on that trial so the overall average size of targets remained unchanged.

The average sizes of both sets were either small (0.7° visual angle) or large (1.3° visual angle), resulting in four experimental conditions: (1) small task-relevant/large task-irrelevant (Fig. 1A), (2) small task-relevant/small task-irrelevant (Fig. 1B), (3) large task-relevant/small task-irrelevant (Fig. 1C), and (4) large task-relevant/large task-irrelevant circles (Fig. 1D). For clarity, conditions (1) and (3) are referred to as negative and positive size-contrast, respectively, corresponding to the expected underestimation and overestimation of the average size of the task-relevant set, whereas conditions (2) and (4) are referred to as size-match.

The experimental conditions were presented block-wise, with each block containing trials from only one condition. The individual sizes of circles in task-relevant and task-irrelevant sets were determined based on a normal cumulative distribution. Specifically, the size of each circle was varied with a standard deviation of 0.15° visual angle around the corresponding mean (0.7° or 1.3° visual angle), and a minimum centre-to-centre distance of 3° visual angle was maintained between adjacent circles. The distance from the fixation cross to the edge of the closest circle was approximately 2.65° visual angle for the small set and 2.35° visual angle for the large set. In this way, we manipulated the heterogeneity of the set on each and every trial while keeping the spatial distribution comparable across trials, as well as across different task-relevant and task-irrelevant object sizes.

The method of constant stimuli was used to detect the perceived average size of the task-relevant circles. The size of the comparison circle varied around the average size of the task-relevant circles with 0.1° visual angle increments, resulting in two different comparison size lists for small (0.7°) and large (1.3°) averages. Ten different comparison sizes were used (half of them were smaller than the average size of the task-relevant set and the other half was larger). In the fMRI session only, one of five fixation durations (500 ms, 1000 ms, 1500 ms,

2000 ms, and 2500 ms) was randomly assigned at the beginning of each trial to introduce temporal jitter (Fig. 1, time window 1).

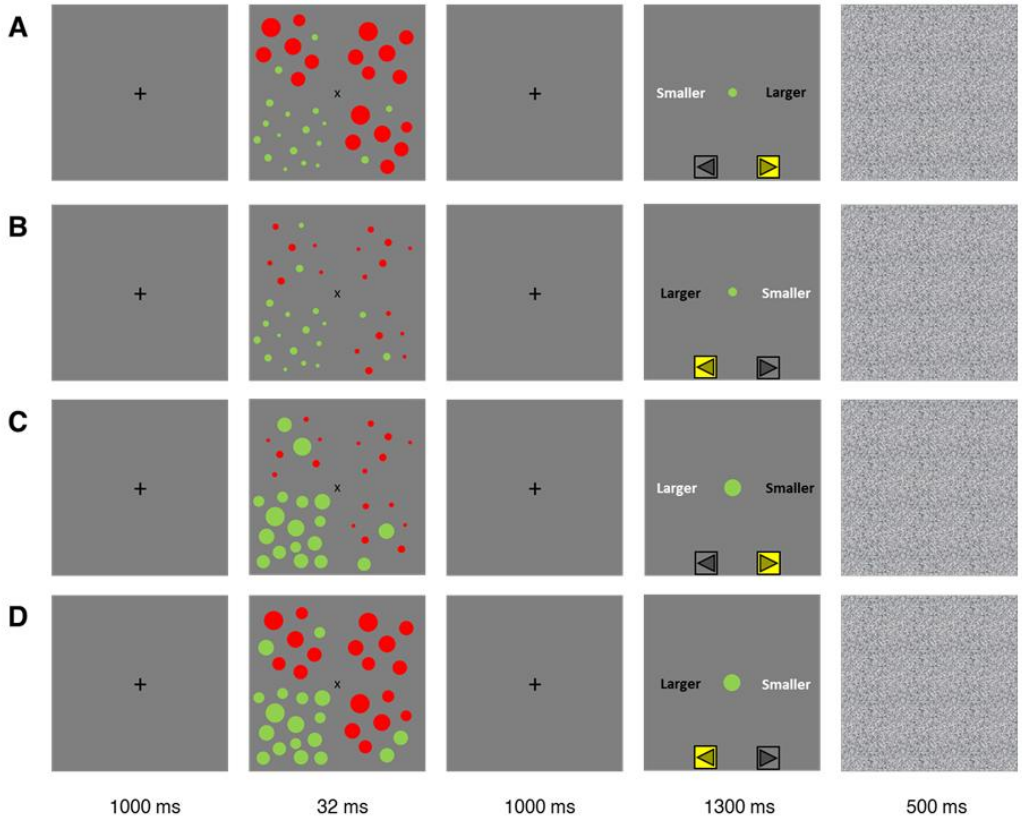

Fig. 1. Illustration of the main task. *The figure shows the experiment setup in which task-relevant items were green, and task-irrelevant items were red. Participants were required to determine whether the comparison circle (Time Window 4) was larger or smaller than the average size of the green circles (Time Window 2). In the size-match conditions (B and D), the average sizes of green and red circles were identical, either small or large, respectively. In the negative size-contrast condition (A), the average size of task-relevant circles was smaller than that of task-irrelevant circles. Conversely, in the positive size-contrast condition (C), the average size of green circles was larger than that of red circles.*

Procedure

Participants attended two sessions: a pre-test session and an MRI session, conducted on different days. Initially, participants attended a pre-test session that lasted approximately 1

hour, during which they completed a colour vision test (Velhagen & Broschmann, 1997), an Ebbinghaus screening task (see Ebbinghaus Screening section), and a practice run for the main task. Participants attended an MRI session 1-4 days after the pre-test session. For some participants the time between the pre-test and the MRI session was longer (from 8 to 27 days). Those participants received an additional 5 minutes of practice before scanning. The MRI session lasted around 1.5 hours and included a position localizer task and the main task performed inside the scanner.

In the main task, participants were instructed to indicate via button press whether the comparison circle was larger or smaller than the average size of the task-relevant set of circles (Fig. 1, time window 4). Response mapping was counterbalanced across trials: in half of the trials, a left button press indicated that the comparison circle was larger, while in the remaining trials, the left button press indicated that the comparison circle was smaller. The response mapping was indicated during the response period (e.g. larger/left, smaller/right).

Each trial started with a fixation period in which only a fixation cross was presented over a grey background (Fig. 1). Following this, task-relevant and task-irrelevant circles appeared in their respective quadrants for 32 ms. After a 1000 ms fixation period, a comparison circle was displayed at the centre of the screen for 1300 ms (Fig. 1). If participants did not respond within the given time frame of 1300 ms, the trial was considered as incorrect. Each trial ended with a grey noise pattern for 500 ms. Participants completed 120 trials per block, resulting in a total of 480 trials (2 task-relevant size × 2 task-irrelevant size × 10 comparison sizes × 4 quadrants × 3 repetitions), which lasted approximately 37 minutes.

Ebbinghaus Screening

During the pre-test session, participants completed a size judgment task designed to assess the strength of the Ebbinghaus illusion, to test whether a potential size-contrast effect

in the main experiment is correlated with an individual's susceptibility to the Ebbinghaus illusion. The task required participants to judge the size of a target circle (0.9° visual angle) that appeared randomly at one of the four corners within an imaginary square (10° visual angle) centred on the fixation cross (Fig. 2). The target circle was green (76.71 cd/m²), and the inducers were red (18.27 cd/m²). Each trial started with a 1000 ms fixation period, followed by a 32 ms presentation of the target circle surrounded by either small (0.7° visual angle) or large (1.3° visual angle) inducers to manipulate the perceived size of the target (Fig. 2A-B). After a 320 ms fixation period, a comparison circle appeared at the centre of the screen, and participants were asked to judge whether it was larger or smaller than the target circle. Participants pressed the right key if the comparison circle was larger and the left key if it was smaller. Each trial ended with a noise screen for 500 ms. Participants performed 80 trials (2 inducer sizes x 4 target locations x 10 comparison sizes) which took approximately 5 minutes to complete.

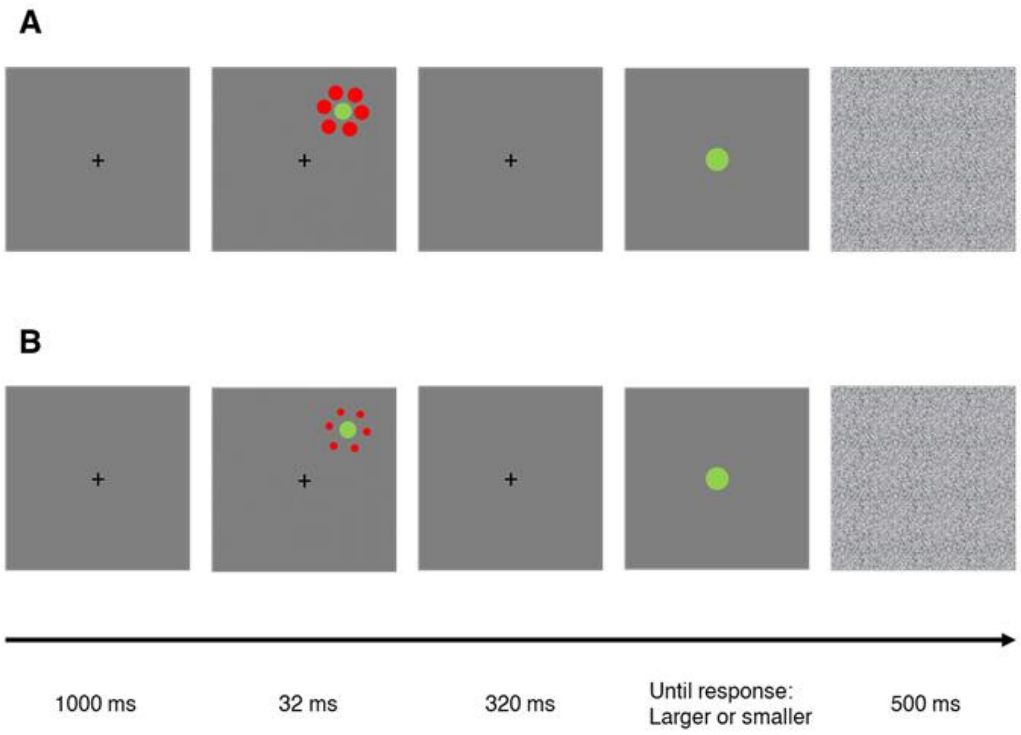

Fig. 2. Illustration of the task in Ebbinghaus screening. *Participants were asked to*

determine whether the comparison circle (time window 4) was larger or smaller than the target circle (time window 2). The target circle was surrounded by either a large (A) or a small inducer (B).

Eye-tracking data acquisition

Eye movement data were recorded from the right eye during the pre-test session using an Eyelink 1000 (SR Research, Mississauga, Ontario, Canada) at a sampling rate of 500 Hz. Participants performed a five-point calibration and validation procedure. A three-times larger circular area with a radius of 1.5° visual angle around the fixation cross was defined as a region of interest (ROI). Eye movement data were recorded throughout the entire experiment, but only critical periods were further analysed (Fig. 1, time window 1-2-3). The average coordinates of the fixation cross for each trial were employed as a drift check for that specific trial. Preprocessing of the eye movement data was performed using RStudio Version 4.2.0 (RStudio Team, 2015).

Behavioural data analysis

The perceived object sizes were calculated and analysed throughout all experimental sessions. To quantify how participants perceived the average size, psychometric curves were generated for each condition by analysing response proportions at 0.1° intervals, reflecting the likelihood of judging the comparison circle as larger than the target circle. A logistic function was used to model this probability (P), and the Point of Subjective Equality (PSE) was calculated as $P = 0.5$, reflecting the size at which the comparison circle was perceived as equal in size to the target circle. A 2 × 2 repeated-measures ANOVA was conducted on PSEs to examine the effects of task-relevant size and task-irrelevant size (small vs. large) in the pre-test and fMRI session.

When fitting the psychometric curves to the data, we calculated the goodness of fit value, quantified as 1 minus the ratio of residual to total variance in response proportions. The obtained curves demonstrated a strong fit in the Ebbinghaus screening task (*r* ranged between .770 and .998), pre-test (*r* ranged between .882 and .997), and fMRI session (*r* ranged between .886 and .994).

To investigate the relationship between participants' susceptibility to the Ebbinghaus illusion and size-contrast effects measured in both the pre-test and fMRI task, we calculated individual size-contrast effects and examined their correlations with Ebbinghaus illusion strength. Size-contrast effects were computed separately for negative and positive size-contrast conditions. Negative size-contrast refers to a condition in which contextual contrast leads to a decrease in perceived size, whereas positive size-contrast indicates an increase in perceived size. These two variants were calculated by using the formulas:

$$Negative\ size-contrast\ effect = PSE(size\_match) - PSE(negative\ size\_contrast)$$

$$Positive\ size-contrast\ effect = PSE(positive\ size\_contrast) - PSE(size\_match)$$

The strength of the Ebbinghaus illusion was quantified as the percentage difference between PSE values obtained with small and large inducers, relative to the target size, using the formula:

$$Illusion\ strength\ (\%) = \frac{(PSE\ small\ inducer - PSE\ large\ inducer) \times 100}{0.9}$$

2.3.3. fMRI measurement

Data acquisition

Each scanning session included both structural (TR = 2500 ms, TE = 2.22 ms, flip angle = 7°, FOV = 240×240 mm, voxel size = 0.94×0.94×0.94 mm, number of slices = 208, slice thickness = 0.94 mm) and functional (TR = 800 ms, TE = 37 ms, flip angle = 52°, FOV = 280×280 mm, voxel size = 2×2×2 mm, number of slices = 72, slice thickness = 2 mm).

MRI acquisitions. Functional magnetic resonance imaging (fMRI) data were acquired to measure blood oxygenation level-dependent (BOLD) signal changes while participants performed the task. Functional scans were acquired using a 3-T PRISMA MRI system (Siemens, Erlangen, Germany) with T2*-weighted EPI sequence. During imaging, visual stimuli were presented via binocular video goggles (NNL, Bergen, Norway) attached to the 64-channel head coil and adjusted to fit the participants' vision. Participants were provided with two 5-button response units (Psychology Software Tools Celeritas, Sharpsburg, PA, USA), and used their index fingers to indicate responses. The button on the left hand corresponded to a "left" response, while the button on the right hand corresponded to a "right" response. In the position localizer task, 450 volumes were acquired, and 2790 volumes were obtained from the main task.

Data preprocessing

The fMRI data were analysed using the statistical parametric mapping software SPM25 (Wellcome Department of Imaging Neuroscience, London; http://fil.ion.ucl.ac.uk/spm/software/spm25). Functional images from both the main experiment and the position localizer task were first realigned to correct for inter-scan movement by aligning each image to the participant's mean functional image. Next, each participant's mean image was normalised to the standard MNI single-subject template using the unified segmentation approach in SPM25. Finally, to improve the signal-to-noise ratio and compensate for minor anatomical differences, the images were smoothed using an 8-mm full width at half-maximum (FWHM) Gaussian kernel.

2.3.4. fMRI analysis

From an initial sample of twenty-nine participants, seven were excluded from all subsequent analyses due to poor task performance and head motion. One participant was

excluded due to a high number of missed responses during the task (more than 10% of task responses, exceeding two standard deviations above the mean across participants). The remaining six participants were excluded due to head motion during scanning. Head motion was quantified using frame-wise displacement (FD), calculated as the sum of absolute differences in translational and rotational motion parameters between consecutive volumes (Power et al., 2012; Ciric et al., 2017). To ensure data quality, a mean FD threshold of 0.2 mm was applied. The final sample consisted of 22 participants (M = 28.55 years, SD = 3.91, 8 females) and was included in all behavioural and functional analyses.

Position Localizer Task

Prior to the fMRI experiment, a position localizer task was conducted to identify the cortical representations in each quadrant of the visual field (lower-left, lower-right, upper-left, and upper-right). This task involved the presentation of a contrast-reversing flickering checkerboard, featuring black and white squares, within each quadrant at a frequency of 8 Hz. The size of each square within the checkerboard was 2.9° visual angle, and each checkerboard was presented for 18 seconds. The entire task lasted approximately 6 minutes.

Four regressors indicated the onsets of 18 seconds visual stimulations, with each regressor corresponding to a different quadrant. The hemodynamic response for each condition was modeled using a canonical HRF and its time derivative, with head movement parameters included as additional regressors in the design matrix.

For every participant, first-level analyses were performed, testing for larger BOLD amplitudes in one condition relative to the remaining three, resulting in condition-specific differential contrasts.

These differential contrasts were then entered into second-level group analysis. One-sample t-tests were performed to assess group-level activation patterns ($p < .001$, whole-brain FWE corrected at the peak voxel level). Functional masks representing retinotopically

distinct stimulus locations were generated based on group-level activation maps computed separately for each quadrant (Fig. 4).

Next, each quadrant-specific group-level activation map from the localizer was intersected with probabilistic maps of early visual areas (V1v, V1d, V2v, V2d, V3v, V3d) provided by Wang et al. (2015), registered in MNI space. This procedure resulted in separate functional ROIs defined by both visual area (e.g., V1) and quadrant representation (e.g., upper-left). These functional ROIs were used in the main task to quantify the number of significantly active voxels in the experimental conditions.

Main task

During the scanning session, participants performed the same size-averaging task as in the pre-test session, in which they indicated whether the comparison circle was larger or smaller than the average size of the task-relevant set, while ignoring a simultaneously presented task-irrelevant set (Fig. 1). Initially, we defined 16 onset regressors, corresponding to the 16 experimental conditions (2 task-relevant sizes: small, large; x 2 task-irrelevant sizes: small, large; x 4: quadrants: lower-left, lower-right, upper-left, upper-right). This way, we could test the effects of our experimental manipulations separately in each quadrant. The hemodynamic response was modeled using a canonical HRF and its time derivative, with head movement parameters included as additional regressors in the design matrix.

ROI-analyses: Separating task-relevant and task-irrelevant sets of objects

In order to investigate the mutual size-contrast effect of task-relevant and task-irrelevant objects, we first needed to disentangle their neural signals. We defined functional ROIs based on localizer scans to test functional activation in voxels representing one specific quadrant at a time (lower-left, lower-right, upper-left, upper-right). This allowed us to examine the neural representation of the task-relevant set of objects in isolation, as well as the

influence of the task-irrelevant set of objects located in a different quadrant, (located diagonally from the task-relevant set). Later, the quadrant specific functional ROIs were further subdivided into different visual areas (V1, V2, V3).

Testing for mutual influences: Size-contrast vs. Size-match

To test mutual size-contrast effects across ensembles, we compared the number of suprathreshold voxels for a target set of objects (e.g., task-relevant objects with a large average size) depending on the average size of the context set (e.g., task-irrelevant objects with a small average size). The average size of the context set could either match the average size if the target set hence constituting a size-match condition (e.g., task-irrelevant objects with a large average size), or it could have a different average size resulting in a size-contrast condition (e.g., task-irrelevant objects with a small average size).

If the average size of the context set affects the neural coding of the target set, we would expect that the number of voxels activated by the target set to differ between size-contrast and the size-match conditions. This effect can be quantified as:

$$\Delta N_{vox} = N_{vox}(size\_contrast) - N_{vox}(size\_match)$$

where:

- $N_{vox}(.)$=number of suprathreshold voxels in the condition indicated in parentheses;
- "size-contrast" = target set paired with context set of a different average size;
- "size-match" = target set paired with context set of the same average size;
- Positive ΔN_{vox} indicates more activated voxels in the size-contrast condition and negative values indicates fewer.

Based on the size-contrast hypothesis, we can make clear directional predictions about ΔN_{vox}. As indicated below in more detail, a positive size-contrast, where a large target set is

presented with a smaller context set, is expected to generate a positive ΔN_{vox}, whereas a negative size-contrast, where a small target set is presented with a large context set is hypothesised to generate a negative ΔN_{vox}.

Large target sets

Pairing a large target set with a small context generates a positive size-contrast and should increase its perceived size, leading to more activated voxels. In that case ΔN_{vox} is expected to be > 0.

For large task-relevant sets:

$$\Delta N_{vox} = N_{vox}(TR_{large}|TI_{small}) - (TR_{large}|TI_{large}) > 0$$

For large task-irrelevant sets:

$$\Delta N_{vox} = N_{vox}(TI_{large}|TR_{small}) - (TI_{large}|TR_{large}) > 0$$

where for all specific-case formulas:

- $N_{vox}(X|Y)$= number of suprathreshold voxels for set X (target) when presented with set Y(context)
- TR = task-relevant set, TI = task-irrelevant set;
- Subscript large/small = average-size category of that set;
- The vertical bar "|" means "presented with"
- ΔN_{vox}=voxel-count difference size-contrast and size-match conditions

Small target sets

Pairing a small target set with a large context should result in a negative size-contrast and should decrease its perceived size, leading to less activated voxels. In that case ΔN_{vox} is expected to be < 0.

For small task-relevant sets, this can be described as:

$$\Delta N_{vox} = N_{vox}(TR_{small}|TI_{large}) - (TR_{small}|TI_{small}) < 0$$

For small task-irrelevant sets, this can be described as:

$$\Delta N_{vox} = N_{vox}\left(TI_{small}\middle|TR_{large}\right) - (TI_{small}|TR_{small}) < 0$$

To summarise, mutual size-contrast interactions between different sets of objects are reflected in differences between size-contrast and size-match conditions. The resulting voxel-count difference ΔN_{vox} is expected to be positive for positive size-contrasts (target larger than context) and negative for negative size-contrasts (target smaller than context).

Functionally, four differential contrasts per quadrant were defined to compare the size-match conditions with the size-contrast conditions (Size-match > negative Size-contrast and positive Size-contrast > Size-match) for task-relevant (Fig. 6A-B) and task-irrelevant (Fig. 6C-D) stimuli.

Statistical testing

By using the functionally defined and intersected ROIs, we extracted the number of significantly active voxels in each functional ROI from the main task differential contrasts, applying a threshold of $p < .05$ (uncorrected), with this voxel count serving as the dependent variable. To control for variability in functional ROI size, the extracted voxel counts were normalised by calculating the percentage of activated voxels relative to the total voxel count within each functional ROI, enabling comparisons across visual areas and quadrants. The normalised percentage of voxels from differential contrasts was submitted to second-level one-sample t-tests to determine whether the observed activations for each contrast were significantly greater than zero, indicating that the functional ROI showed a reliable difference in activation between conditions. In other words, this would be taken as evidence of increased modulation in one condition (e.g., Positive Size-contrast) relative to the other (e.g., Size-match condition). Similarly, greater modulation is expected when comparing the size-match condition with the negative size-contrast.

2.3.5. Behavioural results

Eye movement data

Eye movement data were analysed for the twenty-two participants. We employed a 2 × 2 repeated measures ANOVA to examine fixation maintenance within the fixation ROI across experimental conditions, with factors of task-relevant size (small, large) and task-irrelevant size (small, large). During the critical periods of the experiment, participants' gaze was located within the fixation ROI for an average of 96.70% of the time. The ANOVA did not reveal significant main effects of task-relevant size [small vs. large] (F (1, 21) = 0.53, p = .474, η^2p = 0.025), and task-irrelevant size [small vs. large] (F (1, 21) = 0.34, p = .567, η^2p = 0.016). No significant interaction was observed between the task-relevant and task-irrelevant size of objects (F (1, 21) = 0.22, p = .646, η^2p = 0.010), indicating comparable fixation maintenance across all experimental conditions.

Ebbinghaus Screening

Figure 3 displays the mean PSEs with standard errors for within-subject contrast for the small and large inducer conditions. The Ebbinghaus screening task was conducted to assess participants' susceptibility to size-contrast effects. A two-tailed paired-sample t-test indicated a significant difference in the mean PSE values for small and large inducers (t (21) = 7.51, p < .001, Cohen's d = 1.601).

Specifically, participants estimated the target size as significantly larger in the small inducer condition (M = 0.78, SE = 0.03) than in the large inducer condition (M = 0.66, SE = 0.02), showing that the Ebbinghaus screening task significantly altered the perceived size of the target stimulus. Overall, the illusion strength was approximately 13.77% across participants.

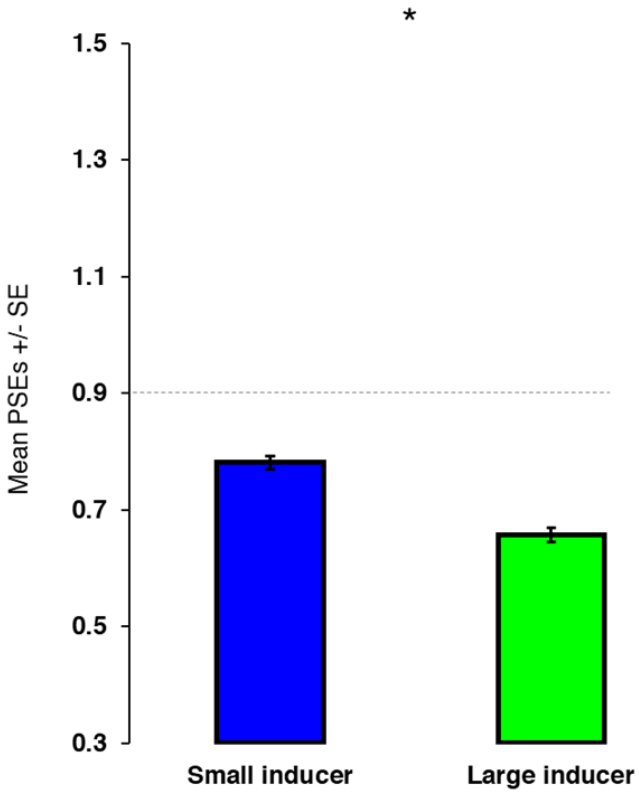

Fig. 3. Perceived target size in Ebbinghaus screening. *Averaged PSEs across different inducer types were plotted. The blue bar represents the small inducer condition, and the green bar indicates the large inducer condition. Asterisks (*) represent significant differences at $p < .05$. Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014). The horizontal dashed grey line represents the physical size of the target stimulus.*

Pre-test

Figure 4A represents the mean PSEs with standard errors for within-subject contrast for the task-relevant size and task-irrelevant size. As expected, participants perceived the average size as significantly smaller in the small task-relevant condition (M = 0.59, SE = 0.02) than in the large task-relevant condition (M = 1.24, SE = 0.02). Additionally, the estimated perceived size was larger in the small task-irrelevant condition (M = 0.97, SE = 0.02) compared to the large task-irrelevant condition (M = 0.86, SE = 0.02). A 2 × 2 ANOVA

on mean PSEs revealed significant main effects of task-relevant size [small vs. large] (F (1, 21) = 1384.40, $p < .001$, $\eta^2p = 0.985$), and task-irrelevant size [small vs. large] (F (1, 21) = 62.88, $p < .001$, $\eta^2p = 0.750$). However, there was no interaction between the two factors (F (1, 21) = 0.16, $p = .698$, $\eta^2p = 0.007$).

To directly assess the size-contrast effect, two-tailed paired-sample t-tests were conducted comparing each contrast condition to the size-match condition (Fig. 4A). Overall, the average size of a set of objects was altered by the average size of another group of objects. In particular, the perceived average size of small task-relevant object sets was significantly smaller when they were presented alongside large task-irrelevant item sets (M = 0.54, SE = 0.02), compared to when they were presented with small task-irrelevant objects (M = 0.64, SE = 0.02), t (21) = 5.06, $p < .001$, Cohen's d = 1.078. Likewise, the perceived average size of large task-relevant object sets was significantly larger when they were presented with small task-irrelevant objects sets (M = 1.29, SE = 0.03) compared to when they were presented alongside large task-irrelevant item sets (M = 1.18, SE = 0.02), t (21) = 5.11, $p < .001$, Cohen's d = 1.088.

Together, these results demonstrate a robust size-contrast effect. Specifically, the perceived average size of the task-relevant set was systematically modulated by the size of the surrounding objects even though they were assigned as task-irrelevant. This finding is consistent with our hypothesis that the average size of a task-irrelevant set influences the perceived average size of the task-relevant set.

fMRI-behaviour

The behavioural results obtained during the fMRI session replicate the pre-test findings, showing that the perceived average size was modulated by the size of task-irrelevant objects (Fig. 4B). As in the pre-test, the perceived average size was smaller in the small task-relevant condition (M = 0.57, SE = 0.02) than in the large task-relevant condition (M = 1.23,

SE = 0.02). Also, participants reported larger perceived average size when task-irrelevant items were small (*M* = 0.93, *SE* = 0.02) compared to when they were large (*M* = 0.87, *SE* = 0.02). Also the 2 x 2 ANOVA on mean PSEs replicated the findings from the pre-tests. It confirmed the significant main effects of both task-relevant size [small vs. large] (F (1, 21) = 1405.43, $p < .001$, $\eta^2 p = 0.985$), and task-irrelevant size [small vs. large] (F (1, 21) = 26.52, $p < .001$, $\eta^2 p = 0.558$). The interaction between the two factors was, consistent with the findings from the pre-tests not significant (F (1, 21) = 0.23, $p = .639$, $\eta^2 p = 0.011$).

Two-tailed paired-sample t-tests confirmed the size-contrast effect, replicating findings from the pre-test (Fig. 4B). The perceived average size was significantly smaller in the small task-relevant and large task-irrelevant condition (*M* = 0.54, *SE* = 0.02) than in the size-match condition (*M* = 0.60, *SE* = 0.02), $t(21) = 3.80$, $p < .001$, Cohen's $d = 0.810$. Similarly, the perceived average size was significantly larger in the large task-relevant and small task-irrelevant condition (*M* = 1.27, *SE* = 0.03) compared to the size-match condition (*M* = 1.20, *SE* = 0.03), $t(21) = 4.50$, $p < .001$, Cohen's $d = 0.960$.

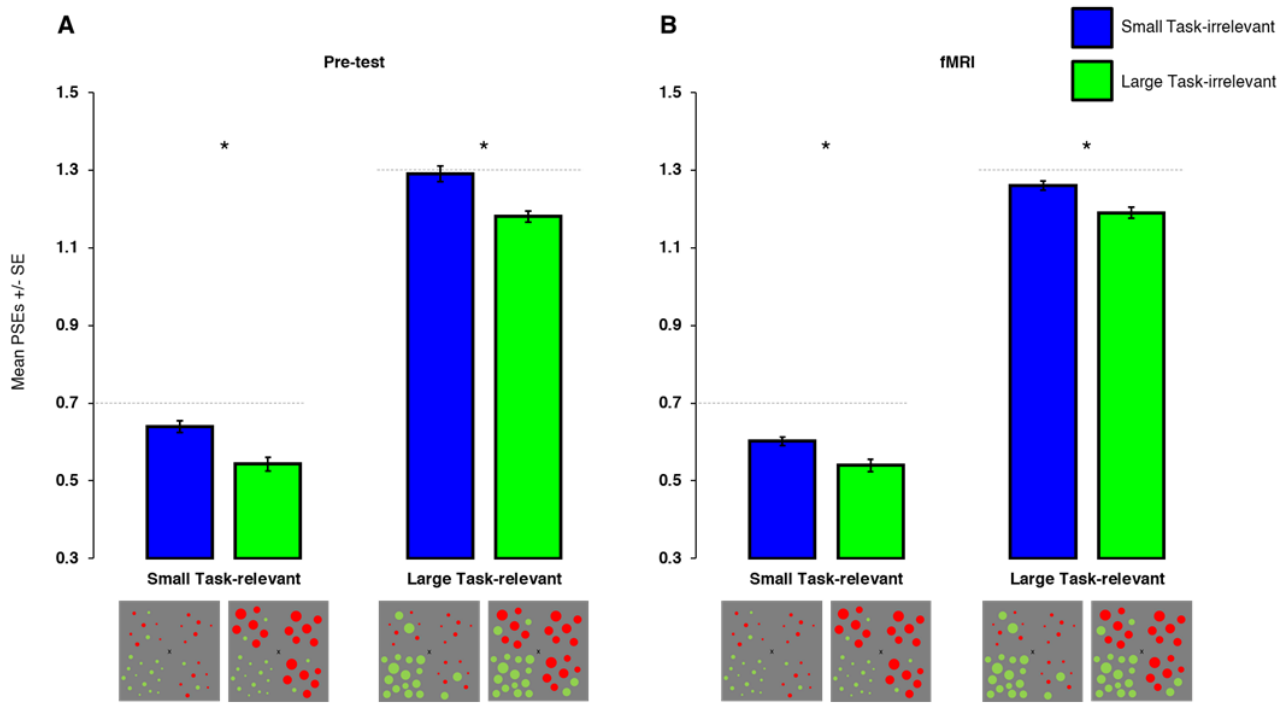

Fig. 4. Perceived average size of experimental conditions in the pre-test (A) and during the fMRI (B). *Averaged PSEs were plotted against the task-relevant size and task-irrelevant size. In each panel, bars on the left reflect the small task-relevant conditions, and bars on the right reflect the large task-relevant conditions. The blue bars represent conditions with small task-irrelevant items, while the green bars represent those with large task-irrelevant items. The horizontal dashed grey lines represent the physical average size of the stimulus display. The figures shown below the x-axis are illustrations of the corresponding experimental conditions. Asterisks (*) represent significant differences at $p < .05$. Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014).*

Correlation analysis

Pearson correlation analyses were conducted to examine the relationship between Ebbinghaus illusion strength and size-contrast effects in both the pre-test and the fMRI session. In the pre-test, no significant correlations were found between illusion strength and negative size-contrast effects (r = –0.206, p = .357), or positive size-contrast effects (r = –0.332, p = .132). Similarly, in the fMRI session, correlations between the illusion strength

and negative (r = 0.118, p = .601) or positive (r = 0.409, p = .06) size-contrast effects were not significant. These findings suggest that participants' susceptibility to the Ebbinghaus illusion is unlikely to be associated with size-contrast effects, implying that size-contrast operates through different mechanisms for ensembles compared to single items, despite producing similar perceptual outcomes.

2.3.6. fMRI results

Position localizer

Figure 5 displays the four quadrant-specific functional ROIs derived from the position localizer task. Whole-brain statistical maps were thresholded at $p < .001$, FWE-corrected at the peak level, to identify significant activation clusters evoked by each quadrant. This analysis revealed distinct peak activations consistent with the retinotopic organisation of the early visual cortex. Specifically, stimuli presented in the lower right quadrant (Fig. 4, blue) elicited activation in the left hemisphere above the calcarine sulcus, while stimuli in the lower left quadrant (Fig. 4, green) activated the right hemisphere above the calcarine sulcus. Likewise, stimuli from the upper right quadrant (Fig. 4, yellow) were represented in the left hemisphere below the calcarine sulcus, and those from the upper left quadrant (Fig. 4, red) were observed in the right hemisphere below the calcarine sulcus. These identified regions were then intersected with probabilistic atlas (e.g., V1, V2, V3) from Wang et al. (2015) to generate functional ROI masks for functional analysis in the main experiment.

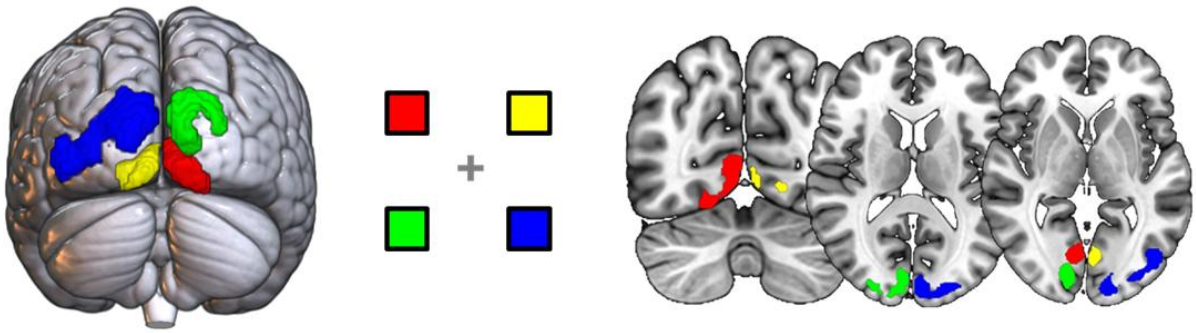

Fig. 5. Visual activation patterns from the position localizer task. *The left panel shows quadrant-specific retinotopic ROIs identified through whole-brain analysis, displayed on a 3D surface-rendered brain. The middle panel illustrates the spatial configuration of the stimulus locations used to define retinotopically distinct regions. The right panel presents axial and coronal slices showing the corresponding peak activations, thresholded at $p < .001$ (FWE-corrected at the peak level).*

Main experiment

Figure 6 illustrates the normalised percentage of activated voxels in early visual areas (V1, V2, V3) for four differential contrasts, averaged across visual field quadrants (lower-right, lower-left, upper-right, upper-left). Each differential contrasts test for effects of a context set on the neural representation of a target set, by contrasting size-contrast conditions with size-match conditions. These differential contrasts were calculated separately for large target sets testing for positive size-contrast effects ($\Delta N_{vox} > 0$) and for small target sets, testing for negative size-contrast effects ($\Delta N_{vox} < 0$). In addition, differential contrasts were calculated separately for task-relevant and task-irrelevant target sets.

Task-relevant sets: small average size

We tested whether a negative size-contrast reduced the number of activated voxels.

For clarity of presentation, voxel-count differences were computed as:

$$\Delta N_{vox} = N_{vox}(size\_match) - N_{vox}(size\_contrast)$$

so that reduced number of activated voxels in the size-contrast condition would generate positive ΔN_{vox}. One-tailed one-sample t-tests against zero were conducted on the percentage of significantly activated voxels extracted from each functional ROI to estimate the negative-size contrast effect (Fig. 6A). Negative size-contrast significantly affected the neural coding of average-size in all functional ROIs [V1 ($t(21) = 3.09$, $p = .003$, Cohen's $d = 0.660$); V2 ($t(21) = 3.06$, $p = .003$, Cohen's $d = 0.653$), and V3 ($t(21) = 3.06$, $p = .003$, Cohen's $d = 0.653$)].

Task-relevant sets: large average size

Similarly, we tested for a significant positive size-contrast effect for task-relevant sets (Fig. 6B). Since a positive size-contrast effect is expected to increase the number of activated voxels in the size-contrast condition, the voxel-count differences were computed as:

$$\Delta N_{vox} = N_{vox}(size\_contrast) - N_{vox}(size\_match)$$

One-tailed one-sample t-tests against zero on the resulting voxel counts revealed significantly more activated voxels ($\Delta N_{vox} > 0$) in the size-contrast condition in all functional ROIs [V1, $t(21) = 3.53$, $p < .001$, Cohen's $d = 0.753$; V2 ($t(21) = 3.86$, $p < .001$, Cohen's $d = 0.822$), and V3 ($t(21) = 4.02$, $p < .001$, Cohen's $d = 0.856$)].

Together these results provide robust evidence that neural populations in early visual areas represent perceived average size differences. In combination with the behavioural evidence of contrast-like modulation in perceived average size, these functional imaging results provide strong support for a size-contrast effect, demonstrating that the perceived average size of task-relevant objects was modulated by the size of the task-irrelevant set.

Task-irrelevant sets: small average size

To explore whether a similar pattern exists for task-irrelevant stimuli, we applied the same analysis for task-irrelevant quadrants. For small task-irrelevant sets we tested whether negative size-contrast reduced the number of activated voxels (Fig. 6C).

Again, for clarity of presentation, voxel-count differences were computed as:

$$\Delta N_{vox} = N_{vox}(size\_match) - N_{vox}(size\_contrast)$$

so that reduced number of activated voxels in the size-contrast condition would generate positive ΔN_{vox}. One-tailed one-sample t-tests against zero were conducted on the percentage of significantly activated voxels extracted from each functional ROI to estimate the negative size-contrast effect. Negative size-contrast significantly affected the neural coding of task-irrelevant average-size in all functional ROIs [V1 $t(21) = 2.76$, $p = .006$, Cohen's $d = 0.588$; V2 ($t(21) = 2.86$, $p = .005$, Cohen's $d = 0.609$), and V3 ($t(21) = 2.78$, $p = .006$, Cohen's $d = 0.592$)].

Task-irrelevant sets: large average size

As for task-relevant sets, we tested whether a positive size-contrast increased the number of activated voxels when contrasting size-contrast and size-match conditions (Fig 6D). Positive size-contrast is expected to increase the number of activated voxels in the size-contrast condition, therefore the voxel-count differences were computed as:

$$\Delta N_{vox} = N_{vox}(size\_contrast) - N_{vox}(size\_match)$$

One-tailed one-sample t-tests against zero on the resulting voxel counts revealed significantly more activated voxels ($\Delta N_{vox} > 0$) in the size-contrast condition in all functional ROIs [V1, $t(21) = 3.01$, $p = .003$, Cohen's $d = 0.640$; V2 ($t(21) = 3.27$, $p = .002$, Cohen's $d = 0.698$), and V3 ($t(21) = 3.41$, $p < .001$, Cohen's $d = 0.726$)].

Although we did not obtain direct behavioural evidence for perceived size modulation in the task-irrelevant ensemble, we found a pattern supporting the notion that task-irrelevant objects may also be subject to size-contrast effects.

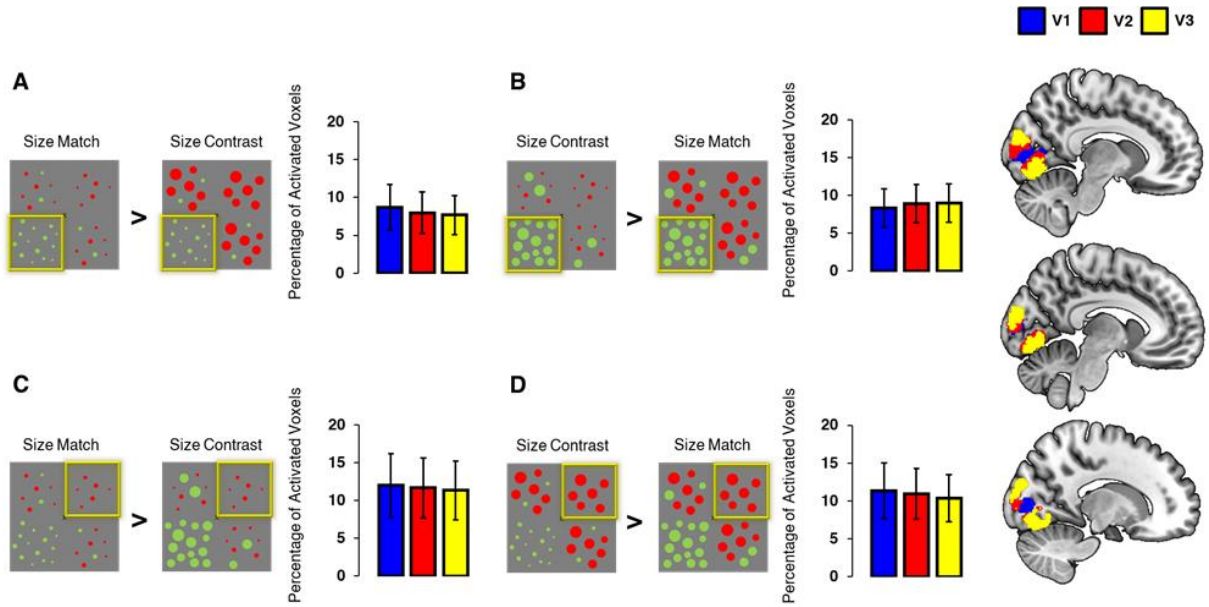

Fig. 6. Normalised percentage of activated voxels per functional ROI. *The figure displays four differential contrasts, with panels A and B showing task-relevant quadrants and panels C and D showing task-irrelevant quadrants. Each panel illustrates a pairwise comparison between the size-match and the size-contrast condition, followed by the percentage of activated voxels within V1 (blue), V2 (red), and V3 (yellow). Error bars indicate the standard errors around the mean for within-subject contrasts (O'Brien & Cousineau, 2014).*

2.3.7. Discussion

Mutual size-contrast in ensemble representations

This study provides strong evidence for a mutual size-contrast effect in ensemble representations, demonstrating that the average size of two groups of objects appears to exert a mutual influence, with each group acting as context for the other. In particular, behavioural data showed that participants perceived the average size of a task-relevant set of objects as larger when presented alongside a smaller set of objects, and smaller when surrounded by a

larger set. This pattern closely replicates classical size-contrast effects, such as those observed in the Ebbinghaus illusion (Ebbinghaus, 1902), but illustrates that this phenomenon exists at the level of ensemble representations. Notably, we observed no significant correlation between the participants' susceptibility to the Ebbinghaus illusion, as detected in the screening task, and the size-contrast effect measured in the main task. This suggests that size-contrast effects may operate via different mechanisms for ensemble summary statistics and individual objects, despite resulting in similar perceptual outcomes.

Retinotopically defined ROI analysis revealed a neural activation pattern indicating that size-contrast modulated the activity evoked by a task-relevant set of objects in early visual areas, in a manner consistent with perceptual size-contrast effects.

For task-irrelevant sets, a similar neural activation pattern was observed, despite the absence of behavioural data confirming a perceived size-contrast effect. The presence of this neural modulation suggests that size-contrast may also have influenced the processing—and potentially the perception—of task-irrelevant objects, even though they were not directly attended or relevant to the behavioural task.

While earlier studies have shown that summary statistics can be computed simultaneously across object groups (Chong & Treisman, 2005), our findings show that these representations are not entirely independent, with size-contrast effects occurring at the level of statistical summary representations.

Our neuroimaging results support the involvement of early visual areas in ensemble size computation, rather than being restricted to higher-level statistical operations (Cant & Xu, 2015; Jia et al., 2022). Although previous studies have shown that perceived, not retinal size, is encoded in early visual areas for individual objects (Murray, Boyaci, & Kersten, 2006; Schwarzkopf, Song, & Rees, 2011; Sperandio, Chouinard, & Goodale, 2012), our findings reveal that such a mechanism applies to ensemble representations, with perceived

average size affecting retinotopic activation. Specifically, conditions in which a small task-relevant set was surrounded by a larger task-irrelevant set showed a reduced number of activated voxels compared to the size-match condition, directly aligning with participants' behavioural responses in which the task-relevant set appeared smaller. Conversely, conditions involving a large task-relevant set surrounded by a smaller task-irrelevant set demonstrated an increased number of activated voxels. These results suggest that the visual system appears to dynamically adjust cortical activation patterns based on perceived size differences in ensemble representations.

Contextual modulation and cross-ensemble dynamics

The present findings align with and extend previous research showing that the visual system automatically rescales ensemble representations relative to contextual cues before statistical summaries are computed (Im & Chong, 2009; Tiurina & Utochkin, 2019; Markov & Tiurina, 2021). Such rescaling implies that ensemble representations reflect perceived rather than retinal size of object groups (Haberman & Suresh, 2021), which is supported by our functionally defined retinotopic analysis. Furthermore, we replicated and broadened our previous findings (Memis et al., 2025), in which we initially observed size-contrast effects where task-irrelevant items influenced the perceived average size of task-relevant items. The current study clarifies these mechanisms, suggesting reciprocal size modulation between task-relevant and task-irrelevant stimuli. Our findings indicate that the visual system constructs ensemble representations of object groups regardless of whether they are attended. Notably, even task-irrelevant groups—those not in the focus of attention—display neural signatures of size-contrast effects, paralleling those seen for attended groups. This implies that ensemble information from both attended and unattended sets are integrated into the scene's relational framework, influencing how other object groups are perceived and represented.

An important question in ensemble perception is whether multiple sets of stimuli are processed independently or combined into a single, pooled representation (Chong & Treisman, 2005). If all stimuli were pooled into a single representation, one would expect an additive averaging effect, in which the presence of large objects increases the overall perceived size of the ensemble. Our results argue against this assumption, instead demonstrating a contrast-like interaction between distinct ensembles. On the contrary, recent work by Ortego and Störmer (2024) found that when participants were asked to report the average orientation of a target-coloured set of lines while ignoring a distractor-coloured set, their judgments were biased toward the distractor orientations, suggesting an integration effect rather than a contrast effect. This biased effect stands in direct opposition to the size-contrast effect we observed for average size judgments. This discrepancy suggests that ensemble perception might be feature-dependent, such as some ensemble properties (e.g., orientation) may be integrated, while others (e.g., size) may be subject to contrast effects. This difference in processing might arise from timing disparities in how features are processed. Previous studies have shown that colour is perceived before orientation (Moutoussis & Zeki, 1997). If similar timing differences exist between size and orientation, this might explain why colour-defined ensembles produce contrast effects for size but result in integration effects for orientation.

Bottom-up and top-down mechanisms in ensemble perception

Our findings raise the question whether size-contrast effects in ensemble representations are primarily driven by early sensory processes or modulated by higher-level feedback mechanisms. The voxel-based changes detected in retinotopic areas may support early sensory origin, aligning with previous behavioural evidence of the rapid extraction of summary statistics, and the robustness of size averaging despite variations in set size and density (Chong & Treisman, 2003; Haberman & Whitney, 2007). While our findings support

the involvement of early retinotopic regions, they do not exclude the possibility of higher-level feedback contribution to these effects. In line with this, previous neuroimaging studies have demonstrated the contribution of higher visual areas. For example, the lateral occipital complex (LOC) has been shown to be involved in perceived size encoding (Weidner & Fink, 2007; Zeng et al., 2020), whereas the right parietal cortex appears to represent illusory size only when it is task-relevant (Plewan et al., 2012). These findings raise the possibility that higher-order regions may contribute to contextual modulation of ensemble statistics.

Taken together, ensemble coding is more likely to be driven by both bottom-up and top-down processes, aligning with the well-known notion of recurrent processing (Lamme & Roelfsema, 2000). Based on this account, retinotopic regions may compute basic ensemble properties, while higher-order cortical regions provide top-down modulation that modifies these representations in response to task demands or contextual factors. This dynamic interplay between the streams allows for iterative refinement of retinal input, potentially occurring simultaneously across multiple ensembles to support comparison of relative properties. Yet, it remains unknown whether these are constructed at the same processing level, or whether a higher-level mechanism extracts and compares these representations.

Although existing neuroimaging studies do not directly examine how multiple ensembles interact, they do show that different brain regions encode distinct ensemble features, supporting the idea that ensemble processing is functionally distributed across the multiple levels of the visual hierarchy. For example, Jia et al. (2022) demonstrated that parieto-occipital areas track global ensemble size rather than individual items, suggesting a top-down mechanism for global integration. Similarly, a study by Cant & Xu (2012, 2015) revealed that the parahippocampal place area (PPA) encodes higher-level ensemble properties such as relative density, while the lateral occipital cortex (LOC) processes local features like shape. However, these studies investigated ensembles defined by a single feature at a time,

leaving open the question of how the visual system handles simultaneously presented but separate ensembles.

Notably, our study was not designed to determine whether the observed size-contrast effects originate solely from early feedforward mechanisms or whether they also involve feedback from higher-level visual areas. Instead, our goal was to test whether retinotopic regions contribute to these effects and provide a functional marker of size-contrast. This approach allowed us to isolate quadrant-specific activations and directly compare activations in task-relevant and task-irrelevant quadrants. A whole-brain analysis was neither feasible nor intended, as it would have introduced confounds from task-irrelevant quadrants and failed to capture the mutual contrast dynamics specific to retinotopically organised areas. Future work could build on these findings by employing DCM to test how early visual areas and higher-level regions interact during ensemble-level size-contrast.

Broader implications for visual perception

Although the current study focuses on ensemble summary statistics, our findings also contribute to the broader discussion of size-contrast illusions, particularly the Ebbinghaus illusion. While some theories attribute the illusion to low-level contour interactions (Chen et al., 2018; Todorović & Jovanović, 2018), others relate the illusion to size-contrast effects (Massaro & Anderson, 1971). However, the observed effects in this study do not reflect the basis of the Ebbinghaus illusion, as the experimental paradigm uses spatially separated ensembles without target-surround structure. Instead, we found that the size-contrast effect extends beyond individual target-inducer interactions to influence ensemble-level statistics. This suggests that similar perceptual mechanisms operate at the level of ensemble statistics, affecting perceived average sizes rather than just individual object sizes.

Taken together, these findings imply that the visual system employs ensemble statistics to segment and compare groups of stimuli. This type of processing may reflect a general mechanism in visual perception. This notion implies that such a mechanism should generalise across different visual dimensions and features, which has to be tested in future experiments, along with the question of how many ensemble representations can be coded at the same time.

The finding that the coding of task-irrelevant objects is altered by context sets of objects in the visual field illustrates that the visual system emphasises differences in ensemble representations of features between groups. This way, not only the appearance or the perceived average size of a group of objects may altered, but it may also be involved in attentional selection. For instance, detecting a pop-out target in visual search may involve ensemble representations. In visual pop-out search, the visual system rapidly computes an average representation of the search array and identifies targets as items that deviate significantly from this statistical norm (Treisman & Gelade, 1980). Therefore, items that strongly differ from the ensemble representations become more salient. The contextual modulation observed in this study indicates that target detection in visual search may be shaped not only by individual stimulus properties but also by the statistical regularities of surrounding context.

Finally, our findings are consistent with predictive coding frameworks, which propose that perception represents the brain's best guess about the most likely interpretation of sensory input given the available context (Rao & Ballard, 1999; Friston, 2005). The visual system does not perceive stimuli in isolation but compares input to internal models based on context, expectations, and prior knowledge. Our findings suggest that the visual system incorporates contextual information from task-irrelevant ensembles to adjust the perceived representation of task-relevant ensembles. The mutual size-contrast effect may thus represent

a manifestation of the visual system's predictive mechanisms, where ensemble representations are continuously updated based on contextual predictions.

2.3.8. Conclusion

Our findings show that the statistical features of simultaneously presented object sets are not computed independently, but are shaped by size-contrast mechanisms. Combining behavioural results with functional imaging analyses, we found that the perceived average size of the task-relevant ensemble was modulated by the size of task-irrelevant objects. Interestingly, a similar pattern was observed for regions representing objects that participants were explicitly instructed to ignore. This mutual size-contrast effect suggests that each ensemble served as a reference for the other, irrespective of task relevance. Together, these findings contribute to a broader understanding of how the visual system organises complex scenes and highlight the importance of size-contrast mechanisms in shaping ensemble representations.

2.3.9. References

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. Vision Research, 101, 34-40.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. Psychological science, 19(4), 392-398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. Psychological science, 12(2), 157-162.

Cant, J. S., & Xu, Y. (2015). The impact of density and ratio on object-ensemble representation in human anterior-medial ventral visual cortex. *Cerebral Cortex*, *25*(11), 4226-4239.

Chen, L., Qiao, C., Wang, Y., & Jiang, Y. (2018). Subconscious processing reveals dissociable contextual modulations of visual size perception. Cognition, 180, 259-267.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. Vision research, 43(4), 393-404.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. Vision research, 45(7), 891-900.

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. Attention, Perception, & Psychophysics, 72, 86-99.

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., ... & Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage, 154, 174-187.

Ebbinghaus, H. (1902). Grundzuge der Psychologie (Vol. 1., Band 2. Theil.). Leipzig: Viet & Co.

Epstein, M. L., & Emmanouil, T. A. (2021). Ensemble statistics can be available before individual item properties: Electroencephalography evidence using the oddball paradigm. Journal of Cognitive Neuroscience, 33(6), 1056-1068.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. Behavior research methods, instruments, & computers, 28(1), 1-11.

Friston, K. (2005). A theory of cortical responses. Philosophical transactions of the Royal Society B: Biological sciences, 360(1456), 815-836.

Haberman, J., & Suresh, S. (2021). Ensemble size judgments account for size constancy. Attention, Perception, & Psychophysics, 83(3), 925-933.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. Current biology, 17(17), R751-R753.

Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. Attention, Perception, & Psychophysics, 71(2), 375-384.

Jia, J., Wang, T., Chen, S., Ding, N., & Fang, F. (2022). Ensemble size perception: Its neural signature and the role of global interaction over individual items. *Neuropsychologia*, *173*, 108290.

Markov, Y. A., & Tiurina, N. A. (2021). Size-distance rescaling in the ensemble representation of range: Study with binocular and monocular cues. Acta Psychologica, 213, 103238.

Massaro, D. W., & Anderson, N. H. (1971). Judgmental model of the Ebbinghaus illusion. Journal of experimental psychology, 89(1), 147.

Memis, E., Yildiz, G. Y., Fink, G. R., & Weidner, R. (2025). Hidden size: Size representations in implicitly coded objects. Cognition, 256, 106041.

Moutoussis, K., & Zeki, S. (1997). Functional segregation and temporal hierarchy of the visual perceptive systems. Proceedings of the Royal Society of London. Series B: Biological Sciences, 264(1387), 1407-1414.

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. Nature neuroscience, 9(3), 429-434.

O'Brien, F., & Cousineau, D. (2014). Representing error bars in within-subject designs in typical software packages. The quantitative methods for psychology, 10(1), 56-67.

Ortego, K., & Störmer, V. S. (2024). Similarity in feature space dictates the efficiency of attentional selection during ensemble processing. Psychonomic Bulletin & Review, 1-10.

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.

Plewan, T., Weidner, R., Eickhoff, S. B., & Fink, G. R. (2012). Ventral and dorsal stream interactions during the perception of the Müller-Lyer illusion: Evidence derived from fMRI and dynamic causal modeling. Journal of Cognitive Neuroscience, 24(10), 2015-2029.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage, 59(3), 2142-2154.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience, 2(1), 79-87.

RStudio Team. (2015). RStudio: Integrated development for R.Boston, MA: RStudio, Inc.

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. Nature neuroscience, 14(1), 28-30.

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. Nature neuroscience, 15(4), 540-542.

Tiurina, N. A., & Utochkin, I. S. (2019). Ensemble perception in depth: Correct size-distance rescaling of multiple objects before averaging. Journal of Experimental Psychology: General, 148(4), 728.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Todorović, D., & Jovanović, L. (2018). Is the Ebbinghaus illusion a size contrast illusion?. Acta psychologica, 185, 180-187.

Velhagen, K., & Broschmann, D. (1997). Tafeln zur Pru¨fung des Farbensinnes. Stuttgart, Germany: Thieme.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. Cerebral cortex, 25(10), 3911-3931.

Weidner , R., & Fink, G. R. (2007). The neural mechanisms underlying the Müller-Lyer illusion and its interaction with visuospatial judgments. Cerebral Cortex, 17(4), 878-884.

Zeng, H., Fink, G. R., & Weidner, R. (2020). Visual size processing in early visual cortex follows lateral occipital cortex involvement. Journal of Neuroscience, 40(22), 4410-4417.

3. General Discussion

In order to interpret the incoming sensory input effectively, the visual system must integrate multiple sources of information from the immediate environment. While it is often assumed that many of these perceptual processes operate automatically, growing evidence suggests that they rely on complex and dynamic mechanisms. As outlined in the theoretical section, previous studies have examined these mechanisms using a range of approaches, including behavioural measures, neuroimaging and neurophysiological techniques, computational models, and predictive coding frameworks. The objective of this thesis is to extend these findings by examining how contextual influences modulate size representations. This chapter investigates the implications and limitations of the studies, and provides suggestions for future research.

3.1. Implications

3.1.1. Visual illusions

Visual illusions robustly illustrate how the visual system interprets the retinal input, showing that two physically identical stimuli can be perceived as differing in size, distance, or lighting. For that reason, they are powerful tools for understanding the mechanism behind visual perception. Study I investigated whether rescaled objects through Ebbinghaus inducers contribute to average size. Results showed that rescaled objects contributed to the average size, regardless of whether they were coded implicitly (masked) or explicitly (non-masked). This suggests that size rescaling takes place relatively early in the visual processing stream, and the rescaled representations become available for further statistical calculation. In a control experiment, we observed a general effect of the Ebbinghaus inducers on the to-be-averaged set. This observation led us to explore whether such contextual modulation also arises from ensemble representations themselves. Namely, study II examined whether the average size of two sets of stimuli affects their perceived size in a contrast-like fashion.

Behavioural and functional data indicated that the perceived average size of the task-relevant set was modulated by the average size of the task-irrelevant set. Also, similar effects were observed for the task-irrelevant set of objects at the functional level. These findings indicate a mutual size contrast effect in simultaneously presented ensembles, suggesting that size contrast operates not only at the item level but also at the level of summary statistics.

Predictive coding theory brings an explanation for how the brain interprets illusory stimuli. According to this framework, perception arises from an interaction between top-down predictions and bottom-up sensory signals, with the brain constantly working to minimise the mismatch between what it expects and what it senses, known as prediction error (Weiss et al., 2002; Friston & Kiebel, 2009; Clark, 2013; Kok & de Lange, 2015). Illusions serve as useful tools for this model, as they reveal the brain's Bayesian inference process, showing how prior knowledge and expectations influence the selection of the most likely perceptual interpretation from ambiguous stimuli. Within this framework, illusions are not seen as perceptual failures but as Bayes-optimal percepts. Specifically, differences between physical and perceived object features are more likely to result from well-known environmental causes (e.g., distance, lighting, or depth), even when the physical stimuli suggest otherwise. The Craik-O'Brien-Cornsweet (CBC) illusion clearly shows how the visual system relies on prior knowledge to interpret sensory data (Purves et al., 2004; Brown & Friston, 2012). In this illusion, two equally bright areas appear different in brightness because of a subtle edge gradient. This happens because the visual system assumes that this gradient signals a change in lighting, rather than a change in surface. As a result, the brain interprets the difference as a variation in lighting across the regions. Likewise, the Ponzo illusion illustrates how prior expectations about size constancy and depth influence perception. When two identical objects are placed between converging lines, the object

located higher appears larger because the visual system interprets it as being farther away in a three-dimensional context.

Another example is perceptual grouping, which plays a critical role in contextual modulation. Specifically, the strength of the Ebbinghaus illusion is highly dependent on the similarity in shape, spatial proximity, and colour between the target and surrounding inducers (Rose & Bressan, 2002). From a predictive coding perspective, such statistical regularities (e.g., grouping by colour) act as priors: the brain assumes that grouped objects belong to a coherent unit and uses this assumption to interpret sensory input (Van de Cruys & Wagemans, 2011). This perspective also explains why the illusion magnitude weakens when grouping principles are violated, potentially resulting in increased prediction error. Supporting this view, Murray et al. (2002) found that activation patterns were influenced by the statistical regularities of visual features. Specifically, they observed increased activity in the LOC and decreased activity in V1 when simple lines were arranged to form a regular configuration, compared to when they were randomly positioned on the screen. These results support predictive coding theory, which suggests that higher-level areas, such as LOC, generate hypotheses about the input and send predictions back to early visual regions. When the input aligns with these predictions, V1 activity is reduced due to the absence of prediction error, reflecting the brain's confidence in the grouped percept. Collectively, these findings reinforce the notion that perception is more likely to be modulated by context and expectations than by merely registering raw sensory input.

3.1.2. Size perception

Although perceiving the size of objects is thought to be an automatic process, the visual system performs complex computations that go far beyond the raw sensory input, integrating it with environmental cues and prior knowledge. Notably, the visual system appears to employ distinct mechanisms for processing individual objects versus groups of

objects (Ariely, 2001). While observers often struggle to report the size of individual items in a set, they are surprisingly accurate at estimating the global properties of the ensemble, such as the average size of the set. Both studies presented in this thesis demonstrate how contextual modulation influences ensemble summary statistics through a series of experiments.

Nevertheless, ensemble coding is not limited to such controlled experimental settings and plays a crucial role in everyday perception by enabling the rapid interpretation of crowded scenes, such as estimating the speed of surrounding traffic, tracking the movement of a crowd, or maintaining a safe distance while driving. To accomplish these goals, the visual system relies on ensemble summary statistics to generate quick and accurate estimates that guide behaviour. Importantly, ensemble coding is not limited to basic stimulus features, such as size or orientation, but also applies to socially meaningful attributes (see Whitney & Yamanashi Leib, 2018, for a review). For instance, observers can extract the average emotion in a group of faces (Haberman & Whitney, 2007), detect the direction of a crowd's movement (Sweeny, Haroz, & Whitney, 2013), or judge the average body size within a group (Oswald, 2023). Together, these findings reinforce the ecological relevance of ensemble coding in everyday life, and demonstrate that investigating size perception within this framework reveals how the visual system efficiently navigates complexity.

In addition to studies with healthy participants, studies involving clinical populations, such as individuals with Autism Spectrum Disorder (ASD), further deepen our understanding of the mechanisms underlying ensemble perception. ASD is a developmental disorder characterised by impairments in both verbal and non-verbal communication, and individuals with ASD tend to prioritise local features over global ones in visual processing (Mottron et al., 1999; O'Riordan et al., 2001; Frith, 2003). This processing bias manifests across several perceptual domains. For instance, individuals with ASD show diminished adaptation effects

in facial identity (Pellicano et al., 2007) and numerosity estimation (Turi et al., 2015), indicating a diminished influence of prior experience on perceptual processing. However, findings related to visual illusions remain mixed. For example, Happé (1996) observed reduced susceptibility to the Ebbinghaus illusion in individuals with ASD, whereas Ropar and Mitchell (2001) found no significant differences between the ASD group and the control group.

Building on these findings, it is reasonable to anticipate that individuals with ASD exhibit deficits in ensemble coding, as this process requires the extraction of global features. However, Corbett and colleagues (2016) found that individuals with ASD exhibited accurate perceptual averaging in mean estimation tasks, suggesting that ensemble mechanisms remain intact despite a bias toward local processing. Taken together, these findings suggest that while certain aspects of contextual integration are impaired in ASD, the ability to extract ensemble-level representations remains intact.

3.1.3. Contextual modulation in ensemble perception

This thesis examines how the visual system integrates contextual information into size representations. Study I showed that contextual modulation occurs prior to the extraction of ensemble representations, as the perceived average size included contributions from implicitly coded (e.g., masked) objects. Study II extended these findings by showing that when two sets of stimuli are presented simultaneously, their average sizes mutually influence each other's perceived size through contrast-like interactions. Together, these findings provide strong support for a processing model (Fig. 3.1.3-1) in which two sets are processed in parallel, with each ensemble being contextually modulated by the other through contrast mechanisms before final ensemble representations are formed. This model aligns with early theories of the Ebbinghaus illusion (Massaro & Anderson, 1971), which propose that surrounding elements serve as standards for size judgments. Notably, even when participants

were instructed to report only the average size of one set of stimuli (e.g., green or red), the visual system appears to automatically compute ensembles from both sets regardless of their task relevance.

These results are consistent with the summary-statistics model, which proposes that ensemble perception involves the rapid and parallel extraction of summary statistics from multiple groups (Ariely, 2001; Chong & Treisman, 2005). However, rather than supporting an independent summary of ensembles, our findings reveal mutual contextual interactions between simultaneously presented ensemble representations. The observed modulation suggests that ensemble processing involves contextual comparison mechanisms that adjust perceived averages based on the relative properties of nearby ensembles. If ensembles were processed independently or if all items were pooled into a single global average, one would predict either no contextual modulation or an additive effect toward the global mean. The robust size contrast effects observed in both studies challenge these assumptions and instead point to a model in which ensembles are processed in parallel but remain contextually interactive. Recent evidence from Ortego and Störmer (2024) provides converging support for the view that ensemble representations are not fully independent. In their study, participants reported the average orientation of differently coloured lines, and their responses were consistently biased toward the task-irrelevant ensemble. Although the direction of modulation differs, they point out that the influence of nearby ensembles is not suppressible. This aligns with our findings, indicating that ensemble representations are processed in parallel but remain susceptible to contextual modulation.

The subsampling model offers an alternative account, proposing that the visual system selectively samples a subset of items rather than computing the average size of the entire set (Myczek & Simons, 2008). According to this framework, the visual system would sample only a few items from each set rather than processing all elements. This strategy would

produce less precise and more variable estimates across trials, as the sampled subset would vary randomly. However, several aspects of our data challenge this explanation. In control conditions where no size contrast was expected (e.g., size-matched conditions), reported average size estimates were remarkably precise and consistent. Such an observation would be less likely within the framework of subsampling, as individual item sizes varied across trials, and random sampling would introduce noise. Thus, subsampling would likely weaken or eliminate the size contrast effect due to inconsistent item selection, contrary to our robust and consistent size contrast findings.

To conclude, the model proposed in this thesis introduces a novel framework by demonstrating that ensemble summary statistics are not mere summaries of stimulus features but instead influence each other's perceived size through contrast-like interactions. This observation suggests that ensemble representations themselves serve as contextual cues in size judgments.

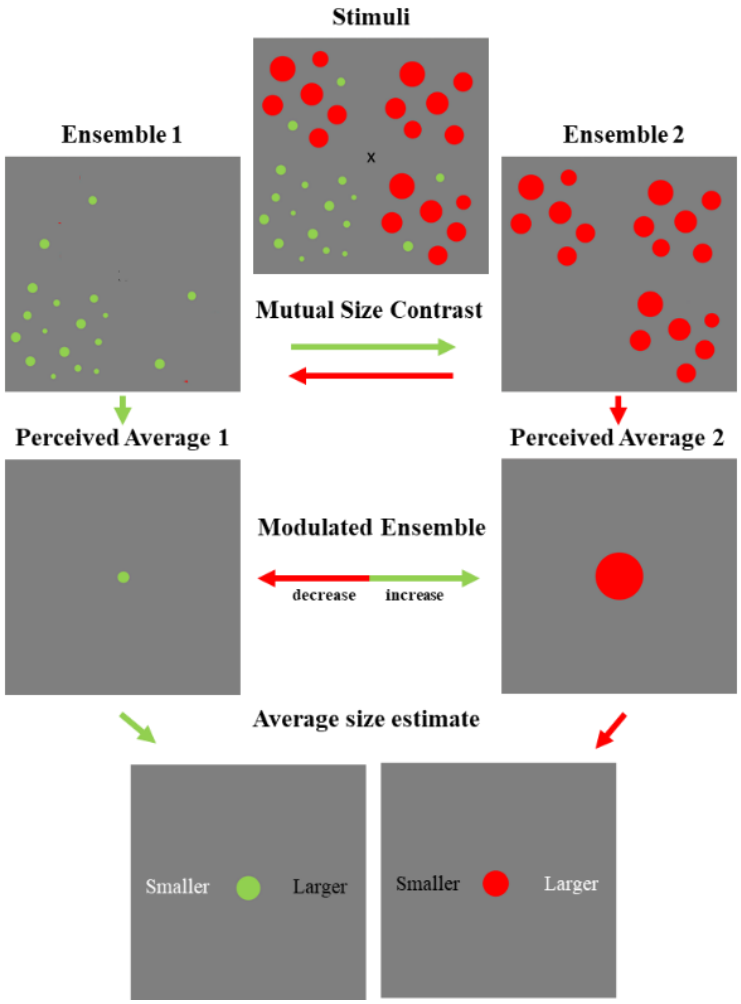

Figure 3.1.3-1 *Schematic illustration of a processing model demonstrating the mutual size contrast effect between two simultaneously presented stimulus sets.*

3.2. Limitations and Future Directions

3.2.1. Study I

Study I examined the level of processing at which ensemble summary statistics are extracted, using a series of behavioural experiments. In Experiment 1, we tested the strength of the Ebbinghaus illusion and observed a significant difference in the perceived target size between the small and large inducer conditions. However, the difference between the control condition (no inducer condition) and the small inducer condition was not significant. One potential explanation is that removing inducers altered the total number of items presented on

the screen. Future studies could easily address this by including a control condition with inducers matched in size to the target, allowing for a more balanced comparison.

Additionally, we applied a 10% illusion strength threshold in Experiment 1 and included only participants who exceeded this criterion in subsequent experiments (Experiments 2, 3, 4A, and 4B). While this controlled for individual variability in illusion sensitivity, future research could investigate these effects without thresholding, to examine a broader range of individual differences. Notably, findings from Study II suggest that contextual modulation in the main task was not solely attributable to the classical Ebbinghaus configuration. Although the target circles within the inducers were removed, the inducers still influenced the perceived average size, suggesting a more general mechanism of contextual modulation.

Furthermore, even though we conducted a control experiment (Experiment 4B) alongside the main task (Experiment 4A), it did not include a matched-size inducer condition, as in Experiment 1, and therefore was not an optimal baseline. One might consider using Experiment 3 as a control condition; however, this is not feasible due to two key differences: (1) the total number of items differed, and (2) individual PSE values were used to define large and small target sizes. The main purpose of Experiment 3 was to test the sensitivity of the paradigm in detecting size changes induced by Experiment 1, rather than to isolate illusion-specific effects.

Our findings demonstrate that rescaled objects contribute to ensemble representations, and this modulation persists even when the objects are implicitly coded (i.e., masked). This suggests that size modulation occurs prior to the level of ensemble formation, while OSM likely disrupts later processing stages. However, since Study I relies solely on behavioural data, it does not provide direct insight into the timing or direction of these effects. Future

research could help clarify whether such contextual modulation arises through feedforward processing alone or also involves feedback from higher-level regions.

3.2.2. Study II

Study II investigated the mutual size contrast effect between simultaneously presented sets of stimuli. The experimental design aimed to isolate the functional activation of both task-relevant and task-irrelevant stimuli within the same trial. To achieve this, stimuli were presented in diagonal quadrants, preserving retinotopic organisation and disentangling neural responses associated with each set. This configuration enabled us to investigate the activation patterns in early visual areas (V1, V2, V3) by using functional ROIs defined by a position localizer task. Due to the spatial specificity of the design, conducting a whole-brain analysis was neither feasible nor intended, as neural responses from surrounding quadrants could have introduced confounding effects. For that reason, the functional analysis was limited to retinotopic regions where the activation can be clearly differentiated based on the quadrants. Nevertheless, future research could employ DCM to examine whether higher-level regions, particularly the LOC, which is known to be involved in illusory size processing (Zeng et al., 2020), modulate early visual areas via feedback signals during ensemble perception.

One methodological limitation was the unequal distribution of stimuli across quadrants. Although there were eighteen green and eighteen red circles in total, the task-relevant quadrant contained fourteen, while the task-irrelevant quadrant contained only six. The remaining quadrants included both task-relevant and task-irrelevant stimuli. This spatial arrangement was necessary to separate task-relevant from task-irrelevant regions while also maintaining distributed attention. Nonetheless, this might have affected the results since we directly compared these two diagonal quadrants.

Additionally, we used size-matched conditions as the control condition, and they were matched in size but not colour. It is possible that different processing may be behind these

features. We had to do it in that way because we wanted to investigate relevant and irrelevant items in the same experimental condition. One can easily test this by using the same-colour control conditions differently for green and red. Even if they are the same size, a colour difference might still create a size-contrast effect.

Another open question is the generalizability of these findings to other visual features. While both studies used size as the ensemble-defining feature, it remains unclear whether the mutual size contrast effect extends to other features, such as orientation, brightness, or spatial frequency. Investigating these dimensions could reveal whether size contrast is feature-specific or a broader ensemble coding mechanism. Lastly, our study demonstrates that the visual system can compute average size across two sets of stimuli. However, it remains unknown whether a third set could also be integrated into this process. Future research could test whether ensemble coding mechanisms simultaneously process three or more sets and whether certain sets receive prioritisation based on grouping cues, such as colour, shape, or spatial configuration.

3.3. Concluding remarks

This thesis provides new insights into the mechanisms underlying ensemble summary statistics. Study I demonstrated that even when the recognisability of objects is significantly impaired by OSM, those objects still contribute to ensemble summary statistics, just as explicitly coded objects do. Notably, this contribution is already rescaled through contextual modulation, in which small Ebbinghaus inducers increased the perceived average size of the to-be-averaged set, while large Ebbinghaus inducers decreased it.

Building on these findings, Study II explored whether these context-driven ensemble representations themselves act as a benchmark that influences the perceived average size of another set. Results showed that the simultaneous presentation of two sets of stimuli produces a contrast effect, meaning that the average size of one set appears larger when the average

size of the other set is smaller. While this effect was robustly observed for the task-relevant items in both behavioural and functional level, retinotopically defined ROI analysis also suggests that task-irrelevant ensembles may be modulated in a similar manner.

Collectively, these findings indicate that ensemble summary statistics go beyond capturing statistical regularities in visual scenes. Rather than merely describing simple object features, they actively influence the perceived size of other objects. The studies presented in this thesis deepen our understanding of size representations in visual perception and provide a strong foundation for future investigation.

4. Summary

4.1. Study I

Perceived object size is not determined solely by the apparent size of the stimulus, but also by the context in which it appears. The Ebbinghaus illusion is an example of how the visual system interprets the target circle as being larger or smaller depending on the size of the surrounding inducers. For instance, the target circle appears larger when surrounded by small Ebbinghaus inducers, and the same object is perceived as smaller when the size of the surrounding inducers is large. In Study I, we investigated the levels of processing at which this interpretation occurs. We combined three distinct psychophysical paradigms into one experiment to investigate whether implicitly coded objects are rescaled through Ebbinghaus inducers before they contribute to the ensemble. To investigate this, six red circles were employed both as inducers of the Ebbinghaus illusion and as masks to initiate OSM. During the main experiment, participants viewed a set of green circles, three of which were surrounded by inducers, and were asked to judge the average size of the green set. Using a method of constant stimuli, we quantified the perceived average size by fitting a logistic function. In total, we tested twenty-nine participants. Three experiments were conducted prior to the main task to test the effectiveness of each paradigm in isolation. Experiment 1 revealed that the perceived size of the target circle was modulated by the size of the inducers, meaning the target circle is perceived as smaller when surrounded by large inducers and larger when surrounded by small inducers. Thus, the Ebbinghaus illusion significantly altered the perceived size of the target. The second experiment aimed to test the strength of OSM in our setup, and we found a significant decrease in the percentage of correct responses in the size discrimination task in the masked conditions compared to non-masked ones. The third experiment showed that the size averaging paradigm is sensitive enough to detect size changes in the magnitude of the ones induced by the Ebbinghaus illusion. We used the

individual PSE values detected in Experiment 1 as the target size. Namely, the perceived average size was larger when target circles were large, and the perceived average size was smaller when target circles were small. Lastly, we combined all paradigms in Experiment 4 to investigate whether implicitly coded objects contribute to ensemble summary statistics. We observed that both masked and non-masked stimuli contributed to average size estimates. Notably, this contribution was already rescaled by the Ebbinghaus inducers. Additionally, we conducted a control experiment in which we removed all target circles within the inducers and tested the general effects of Ebbinghaus inducers on the to-be-averaged display. Interestingly, we observed a size-contrast effect, meaning that small inducers increased the perceived average size, while large inducers decreased it. Taken together, Study I demonstrates that contextual integration via Ebbinghaus inducers alters size representations at an early stage, and this effect is independent of whether the object is implicitly or explicitly coded.

4.2. Study II

The visual system employs various strategies to handle its limited capacity. One such strategy involves calculating the global properties of a scene instead of processing each individual item, a process known as ensemble summary statistics. Study I revealed a size contrast effect in a control experiment testing the general effect of Ebbinghaus inducers, suggesting that summary statistics do not purely reflect simple descriptions of object features, and instead themselves influence the perceived average size of another set. This finding aligns with the early assumption of the Ebbinghaus illusion, which suggests that the illusion arises as a size contrast mechanism, and inducers act as standards for size judgments (Massaro & Anderson, 1971). Study II aimed to investigate whether the average size of two sets of stimuli generates a size contrast effect via presenting two ensembles grouped by colour. During the task, participants responded to the comparison item either as smaller or

larger than the average size of the task-relevant set. A total of twenty-nine participants were tested. To investigate the mutual size contrast effect, task-relevant and task-irrelevant quadrants were placed diagonally, allowing for the observation of activation patterns in line with the retinotopic organisation of the visual cortex. In order to achieve this, four distinct quadrants were functionally defined in the position localizer task, and then used as functional ROIs in the main task to investigate activation patterns in the task-relevant and task-irrelevant quadrants. Behavioural data provided strong support for the hypothesis that the size of task-irrelevant objects modulates the perceived average size of task-relevant ones. Specifically, the perceived average size of the task-relevant set increased when the average size of the task-irrelevant set was small, whereas it decreased when the average size of the task-irrelevant set was large. This perceptual modulation was further supported by systematic changes in functional activation patterns, consistent with the retinotopic organisation. Notably, a similar modulation was observed in regions, representing the task-irrelevant set of objects, suggesting that even the task-irrelevant ensemble may be subject to size contrast effects. These results demonstrate that simultaneously presented ensembles are not coded independently but instead are processed in a contrast-like fashion. All in all, these findings indicate that size contrast effects observed in many illusions, such as the Ebbinghaus illusion, occur at the level of statistical summary statistics.

5. References

Ajina, S., & Bridge, H. (2019). Subcortical pathways to extrastriate visual cortex underlie residual vision following bilateral damage to V1. Neuropsychologia, 128, 140-149.

Aleci, C., & Dutto, K. (2024). Seeing the invisible: theory and evidence of blindsight. Discover Medicine, 1(1), 1-22.

Anderson, J. C., & Martin, K. A. (2009). The synaptic connections between cortical areas V1 and V2 in macaque monkey. Journal of Neuroscience, 29(36), 11283-11293.

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. Nature neuroscience, 10(10), 1313-1321.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological science*, *12*(2), 157-162.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, *61*(3), 183.

Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., ... & Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. Neuron, 29(2), 529-535.

Boring, E. G. (1940). Size constancy and Emmert's law. *The American Journal of Psychology*, *53*(2), 293-295.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. Annu. Rev. Neurosci., 28(1), 157-189.

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of neuroscience*, *17*(6), 2112-2127.

Boycott, B., & Wässle, H. (1999). Parallel processing in the mammalian retina: the Proctor Lecture. Investigative Ophthalmology & Visual Science, 40(7), 1313-1327.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological science*, *22*(3), 384-392.

Brown, H. R., & Friston, K. J. (2012). Dynamic causal modelling of precision and synaptic gain in visual perception—an EEG study. Neuroimage, 63(1), 223-231.

Bugmann, G., & Taylor, J. G. (2005). A model of visual backward masking. *Biosystems*, *79*(1-3), 151-158.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. Annual review of neuroscience, 21(1), 47-74.

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *Journal of Neuroscience*, *32*(22), 7685-7700.

Cant, J. S., & Xu, Y. (2015). The impact of density and ratio on object-ensemble representation in human anterior-medial ventral visual cortex. *Cerebral Cortex*, *25*(11), 4226-4239.

Carlson, T. A., Rauschenberger, R., & Verstraten, F. A. (2007). No representation without awareness in the lateral occipital cortex. *Psychological Science*, *18*(4), 298-302.

Chen, L., Qiao, C., Wang, Y., & Jiang, Y. (2018). Subconscious processing reveals dissociable contextual modulations of visual size perception. Cognition, 180, 259-267.

Chen, L., Wu, B., Yu, H., & Sperandio, I. (2024). Network dynamics underlying alterations in apparent object size. *Brain Communications*, *6*(1), fcae006.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research*, *43*(4), 393-404.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision research*, *45*(7), 891-900.

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. *Attention, Perception, & Psychophysics*, *72*, 86-99.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and brain sciences, 36(3), 181-204.

Corbett, J. E., Venuti, P., & Melcher, D. (2016). Perceptual averaging in individuals with autism spectrum disorder. Frontiers in Psychology, 7, 1735.

Czigler, I., Balázs, L., & Pató, L. G. (2004). Visual change detection: event-related potentials are dependent on stimulus location in humans. *Neuroscience letters*, *364*(3), 149-153.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision research*, *37*(22), 3181-3192.

De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, *108*(32), 13341-13346.

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. Neuron, 70(2), 200-227.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. Journal of Experimental Psychology: General, 129(4), 481.

Ebbinghaus, H. (1902). Grundzuge der Psychologie (Vol. 1., Band 2. Theil.). Leipzig: Viet & Co.

Edelman, G.M. (1992) Bright Air, Brilliant Fire. On the Matter of Mind, Basic Books, USA

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision research*, *44*(12), 1321-1331.

Epstein, M. L., & Emmanouil, T. A. (2021). Ensemble statistics can be available before individual item properties: Electroencephalography evidence using the oddball paradigm. *Journal of Cognitive Neuroscience*, *33*(6), 1056-1068.

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. Nature neuroscience, 8(10), 1380-1385.

Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. *Current biology*, *18*(21), 1707-1712.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience*, *4*, 215.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex (New York, NY: 1991), 1(1), 1-47.

Fischer, D. J., Schröer, F., Denecke, S., Murphy, L., & Kühn, S. (2024). Are we afraid of the woods?–An investigation of the implicit and explicit fear reactions to forests. Environmental Research, 260, 119573.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. Philosophical transactions of the Royal Society B: Biological sciences, 364(1521), 1211-1221.

Frith, U. (2003). *Autism: Explaining the enigma*. Blackwell publishing.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical neurophysiology*, *120*(3), 453-463.

Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychological review*, *58*(6), 460.

Goettker, A., & Stewart, E. E. (2022). Serial dependence for oculomotor control depends on early sensory signals. *Current Biology*, *32*(13), 2956-2961.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. Trends in neurosciences, 15(1), 20-25.

Goodale, M. A., Westwood, D. A., & Milner, A. D. (2004). Two distinct modes of control for object- directed action. Progress in brain research, 144, 131-144.

Goodhew, S. C., Pratt, J., Dux, P. E., & Ferber, S. (2013). Substituting objects from consciousness: A review of object substitution masking. *Psychonomic bulletin & review*, *20*, 859-877.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. Journal of Experimental Psychology: Learning, memory, and cognition, 11(3), 501.

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature*, *199*(4894), 678-680.

Gregory, R. L. (1968). Perceptual illusions and brain models. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *171*(1024), 279-296.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. Annu. Rev. Neurosci., 27(1), 649-677.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current biology*, *17*(17), R751-R753.

Happé, F. G. (1996). Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. A research note. *Journal of child psychology and psychiatry*, *37*(7), 873-877.

Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *The American journal of psychology*, *54*(1), 21-37.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology, 160(1), 106.

Hutchinson, B.T., 2019. Toward a theory of consciousness: a review of the neural correlates of inattentional blindness. Neurosci. Biobehav. Rev. 104, 87–99. doi:10.1016/j.neubiorev.2019.06.003.

Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception, & Psychophysics*, *71*(2), 375-384.

Jia, J., Wang, T., Chen, S., Ding, N., & Fang, F. (2022). Ensemble size perception: Its neural signature and the role of global interaction over individual items. *Neuropsychologia*, *173*, 108290.

Julian, J. B., Ryan, J., & Epstein, R. A. (2017). Coding of object size and object category in human visual cortex. *Cerebral cortex*, *27*(6), 3095-3109.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. Brain research bulletin, 44(2), 107-112.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. Nature reviews neuroscience, 1(2), 91-100.

Kaufman, L., & Kaufman, J. H. (2000). Explaining the moon illusion. *Proceedings of the National Academy of Sciences*, *97*(1), 500-505.

Kilpatrick, F. P., & Ittelson, W. H. (1953). The size-distance invariance hypothesis. *Psychological review*, *60*(4), 223.

Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. An introduction to model-based cognitive neuroscience, 221-244.

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral cortex*, *22*(9), 2197-2206.

Kolb, H. (2003). How the retina works: Much of the construction of an image takes place in the retina itself through the use of specialized neural circuits. American scientist, 91(1), 28-35.

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, *33*(25), 10235-10242.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114-1124.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1481), 857-875.

Lamme, V. A. (2003). Why visual attention and awareness are different. Trends in cognitive sciences, 7(1), 12-18.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends in neurosciences, 23(11), 571-579.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434-1448.

Leibowitz, H., Brislin, R., Perlmutrer, L., & Hennessy, R. (1969). Ponzo perspective illusion as a manifestation of space perception. *Science*, *166*(3909), 1174-1176.

Libedinsky, C., & Livingstone, M. (2011). Role of prefrontal cortex in conscious visual perception. *Journal of Neuroscience*, *31*(1), 64-69.

MacLean, M. W., Hadid, V., Spreng, R. N., & Lepore, F. (2023). Revealing robust neural correlates of conscious and unconscious visual processing: activation likelihood estimation meta-analyses. Neuroimage, 273, 120088.

Markov, Y. A., & Tiurina, N. A. (2021). Size-distance rescaling in the ensemble representation of range: Study with binocular and monocular cues. *Acta Psychologica*, *213*, 103238.

Massaro, D. W., & Anderson, N. H. (1971). Judgmental model of the Ebbinghaus illusion. Journal of experimental psychology, 89(1), 147.

Merigan, W. H., & Maunsell, J. H. (1993). How parallel are the primate visual pathways?. Annual review of neuroscience.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behavioural brain research, 6(1), 57-77.

Mottron, L., Belleville, S., & Ménard, E. (1999). Local bias in autistic subjects as evidenced by graphic tasks: Perceptual hierarchization or working memory deficit?. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(5), 743-755.

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature neuroscience*, *9*(3), 429-434.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. Proceedings of the National Academy of Sciences, 99(23), 15164-15169.

Murray, S. O., Schrater, P., & Kersten, D. (2004). Perceptual grouping and the interactions between visual cortical areas. *Neural Networks*, *17*(5-6), 695-705.

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & psychophysics*, *70*, 772-788.

Nassi, J. J., & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. Nature reviews neuroscience, 10(5), 360-372.

O'riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 719.

Ortego, K., & Störmer, V. S. (2024). Similarity in feature space dictates the efficiency of attentional selection during ensemble processing. Psychonomic Bulletin & Review, 1-10.

Oswald, F., Griffin, J. W., Weisbuch, M., & Adams Jr, R. B. (2023). People watching: Social perception and the ensemble coding of bodies. *Journal of Nonverbal Behavior*, *47*(4), 545-568.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. Science, 292(5516), 510-512.

Pellicano, E., Jeffery, L., Burr, D., & Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Current Biology*, *17*(17), 1508-1512.

Plewan, T., Weidner, R., Eickhoff, S. B., & Fink, G. R. (2012). Ventral and dorsal stream interactions during the perception of the Müller-Lyer illusion: Evidence derived from fMRI and dynamic causal modeling. *Journal of Cognitive Neuroscience*, *24*(10), 2015-2029.

Ponzo, M. (1911). Intorno ad alcune illusioni nel campo delle sensazioni tattili, sull'illusione di Aristotele e fenomeni analoghi [On some illusions in the field of tactile sensations on the illusion of Aristotle and similar phenomena]. Auchive für die Gesamte Psychologie, 16, 307–345.

Purves, D., Williams, S. M., Nundy, S., & Lotto, R. B. (2004). Perceiving the intensity of light. Psychological review, 111(1), 142.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87.

Reiss, J. E., & Hoffman, J. E. (2007). Disruption of early face recognition processes by object substitution masking. *Visual Cognition*, *15*(7), 789-798.

Ropar, D., & Mitchell, P. (2001). Susceptibility to illusions and performance on visuospatial tasks in individuals with autism. *Journal of child Psychology and Psychiatry*, *42*(4), 539-549.

Rose, D., & Bressan, P. (2002). Going round in circles: shape effects in the Ebbinghaus illusion. Spatial Vision, 15(2), 191-204.

Ross, H., & Plug, C. (2002). *The mystery of the moon illusion: Exploring size perception*. Oxford University Press.

Schaeffel, F. (2006). Processing of information in the human visual system. Handbook of machine vision, 1-33.

Schall, J. D. (2015). Visuomotor functions in the frontal lobe. *Annual Review of Vision Science*, *1*(1), 469-498.

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. Nature neuroscience, 14(1), 28-30.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. perception, 28(9), 1059-1074.

Song, C., Schwarzkopf, D. S., & Rees, G. (2011). Interocular induction of illusory size perception. *BMC neuroscience*, *12*, 1-9.

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nature neuroscience*, *15*(4), 540-542.

Stefanics, G., Heinzle, J., Horváth, A. A., & Stephan, K. E. (2018). Visual mismatch and predictive coding: a computational single-trial ERP study. *Journal of Neuroscience*, *38*(16), 4020-4030.

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 329.

Tal, A., Sar-Shalom, M., Krawitz, T., Biderman, D., & Mudrik, L. (2024). Awareness is needed for contextual effects in ambiguous object recognition. Cortex, 173, 49-60.

Tiurina, N. A., & Utochkin, I. S. (2019). Ensemble perception in depth: Correct size-distance rescaling of multiple objects before averaging. *Journal of Experimental Psychology: General*, *148*(4), 728.

Todorović, D., & Jovanović, L. (2018). Is the Ebbinghaus illusion a size contrast illusion?. Acta psychologica, 185, 180-187.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. Neuron, 21(4), 753-759.

Turi, M., Burr, D. C., Igliozzi, R., Aagten-Murphy, D., Muratori, F., & Pellicano, E. (2015). Children with autism spectrum disorder show reduced adaptation to number. *Proceedings of the National Academy of Sciences*, *112*(25), 7868-7872.

Vandenbroucke, A. R., Fahrenfort, J. J., Sligte, I. G., & Lamme, V. A. (2014). Seeing without knowing: neural signatures of perceptual inference in the absence of report. Journal of cognitive neuroscience, 26(5), 955-969.

Van de Cruys, S., & Wagemans, J. (2011). Gestalts as predictions: Some reflections and an application to art. Gestalt Theory, 33(3), 325-344.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. Neuron, 56(2), 366-383.

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision research*, *32*(5), 931-941.

Weidner, R., & Fink, G. R. (2007). The neural mechanisms underlying the Müller-Lyer illusion and its interaction with visuospatial judgments. *Cerebral Cortex*, *17*(4), 878-884.

Weidner, R., Plewan, T., Chen, Q., Buchner, A., Weiss, P. H., & Fink, G. R. (2014). The moon illusion and size–distance scaling—evidence for shared neural patterns. *Journal of cognitive neuroscience*, *26*(8), 1871-1882.

Weidner, R., Shah, N. J., & Fink, G. R. (2006). The neural basis of perceptual hypothesis generation and testing. *Journal of Cognitive Neuroscience*, *18*(2), 258-266.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. Nature Neuroscience, 5(6), 598-604.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of psychology*, *69*(1), 105-129.

Wu, B., Feng, B., Han, X., Chen, L., & Luo, W. (2023). Intrinsic excitability of human right parietal cortex shapes the experienced visual size illusions. *Cerebral Cortex*, *33*(10), 6345-6353.

Wurtz, R. H., & Kandel, E. R. (2000). Central visual pathways. Principles of neural science, 4, 523- 545.

Zeki, S. (1993). A vision of the brain. Blackwell scientific publications.

Zeng, H., Fink, G. R., & Weidner, R. (2020). Visual size processing in early visual cortex follows lateral occipital cortex involvement. *Journal of Neuroscience*, *40*(22), 4410-4417.

Zipser, K., Lamme, V. A., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. Journal of Neuroscience, 16(22), 7376-7389.

6. List of abbreviations

DCM	dynamic causal modeling
EEG	electroencephalography
ERP	event-related potential
fMRI	functional MRI
IPS	intraparietal sulcus
LGN	lateral geniculate nucleus
LOC	lateral occipital cortex
MRI	magnetic resonance imaging
MVPA	multivoxel pattern analysis
OSM	object-substitution masking
PPA	parahippocampal place area
PSE	point of subjective equality
ROI	region of interest
TMS	transcranial magnetic stimulation

7. Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, PD Dr. Ralph Weidner, for his invaluable guidance, expertise, and support at every step of my Ph.D. study. I feel incredibly privileged to work with him, and he has been the best supervisor, always being friendly, encouraging, and understanding. His door was always open, no matter how busy he was, and he always took the time to listen and provide feedback on all my questions. He created an environment where I felt safe to explore and learn from my mistakes, always focusing on solutions and supporting my growth as a researcher. His confidence in my abilities, even when I doubted myself, motivated me to persevere through challenges. His supervision style offered me clarity, confidence, and independence that were essential for my academic progress. I am extremely grateful to have worked under his guidance, and thanks to him, my Ph.D. experience has been more than I could have hoped for. The best part is that I will continue to work alongside him in the future.

I would also like to extend my appreciation to both the former and current members of the Institute of Neuroscience and Medicine (INM-3) at FZJ for their support throughout my studies. Specifically, I would like to thank Prof. Dr. Simone Vossel, whose invaluable support and help were crucial to my progress. I am also grateful to Dr. Gizem Y. Yildiz for her guidance and expertise throughout my work. Furthermore, I would like to thank Prof. Dr. Gereon Rudolf Fink for giving me the opportunity to work at FZJ and for providing valuable feedback at all phases of my research. I would like to thank all my colleagues at FZJ for their helpful discussions and suggestions during the CNS and Brainfast meetings, which helped me refine my ideas and see things from different perspectives. Additionally, I would like to thank Magda and Susanne for their daily chats and continuous administrative support - our brief conversations were always such a pleasure.

A special thanks goes to Assoc. Prof. Dr. Ahu Gökçe, my undergraduate supervisor, who first opened the doors to the career I pursue today. She provided me with the solid foundation upon which all my academic work has been built. Her guidance and encouragement during my undergraduate studies had a lasting impact on my career, and I will always be thankful for her confidence in me.

To my family, I am deeply grateful. I would like to thank my late father, Engin Memiş, for the love and guidance he gave me; his wisdom and encouragement will always stay with me. To my mother, Melek Memiş, my greatest source of strength and my best friend, without whom none of this would have been possible, thank you for your endless love, for being by my side at every step, and for making countless trips just to support me during this journey. To my dear brother, Emir Çetin Memiş, who has been more than just a sibling to me—more like a colleague—I want to sincerely thank you for your constant support and for sharing in both the hard work and the happiness of this achievement. Your presence has meant so much to me throughout this journey. I would also like to express my warmest thanks to my husband, Ayberk Yalçın, whose patience, love, and understanding have always kept me motivated. Thank you for always being my first and the most enthusiastic participant.

Last but not least, huge thanks to my dear friends I met in Germany - Ulises, Shiva, Eren, Valerie, Lea, Nan, Fatemeh, Gianna, Linnea, Gizem, Marina, Shirley, Sinah, Annika, Ivan, Mahan, Franziska, Giovanni, Michella. Thank you all for your friendship, endless encouragement, and for celebrating every milestone with me, as well as for making Germany a home away from home. I couldn’t have done it without you, and I am so grateful to each and every one of you.

8. Curriculum Vitae

Personal Details

Name	Memis, Elif
Date of Birth	1994/10/20
Nationality	Turkish
Gender	Female
Adresse	Marienstraße 2, 52457, Aldenhoven
E-Mail	e.memis@fz-juelich.de

Education

2021-present	Research Center Juelich (Juelich, Germany), Ph.D, Cognitive Neuroscience
2018-2021	Yeditepe University (Istanbul, Turkey) Master of Arts in Cognitive Science, 3.86/4.00
2013-2018	Kadir Has University (Istanbul, Turkey) Bachelor of Psychology, GPA: 3.29/4.00

Academic Activities

05/2018	Poster presentation at the International Meeting of the Psychonomic Society in Amsterdam
04/2019	Poster presentation at International Symposium on Brain and Cognitive Science at Yeditepe University, in Istanbul, Turkey
05/2020	Poster presentation at International Symposium on Brain and Cognitive Science at Bilkent University, in Ankara, Turkey
08/2021	Poster presentation at 43rd European Conference on Visual Perception (online)
08/2022	Good scientific practice workshop in Research Center Juelich, Germany
09/2022	Scientific writing workshop in Research Center Juelich, Germany
09/2022	Scientific presentation workshop in Research Center Juelich, Germany
09/2022	Management of a doctoral project workshop in Research Center Juelich, Germany

04/2023	Innovation and entrepreneurship workshop in Research Center Juelich, Germany
01/2025	Poster presentation at 43rd European Conference on Cognitive Neuropsychology, in Bressanone, Italy

Scholarships and Awards

2017	Kadir Has University 100% honour scholarship (2017)

Publications

Ozdemir, D., **Memis**, E., Güney, Ç., Oskanbaş, A., Sametoğlu, S., Eroğlu, M. G., & İsarlı, Ç. (2018). Change Detection Performance in Location and Object Change. Koç University Undergraduate Psychology Journal, (4), 1-11.

Memis, E., Güney, Ç., Gökçe, A. (2020). Emotional Word Processing Within Configural Representations. Studies in Psychology, 40(1).

Memis, E., Yildiz, G. Y., Fink, G. R., & Weidner, R. (2025). Hidden size: Size representations in implicitly coded objects. Cognition, 256, 106041.

Memis, E., Fink, G. R., & Weidner, R. (*Unpublished manuscript, included in this dissertation and made publicly available via KUPS*). Cross-ensemble size contrast in summary statistics: Neural and behavioural evidence.

Technical Skills

Python, PsychoPy, R, fMRI, MATLAB, SPSS, SPM, MS Office Programs

Languages

Native	Turkish
Proficient	English
Beginner	German

Juelich, den 08.2025