EFFECTS OF CONTEXT AND EXPECTATIONS ON DYNAMICS OF VISUAL PROCESSING

A DISSERTATION SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
NEUROSCIENCE

By
Amna Malik
September 2023

Effects of context and expectations on dynamics of visual processing

By Amna Malik

September 2023

We certify that we have read this dissertation and that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Hüseyin Boyacı(Advisor)

Burcu Ayşen Ürgen

Zahide Pamir Karatok

Didem Kadıhasanoğlu

Buse Merve Ürgen

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

EFFECTS OF CONTEXT AND EXPECTATIONS ON DYNAMICS OF VISUAL PROCESSING

Amna Malik
Ph.D. in Neuroscience
Advisor: Hüseyin Boyacı
September 2023

Living organisms rely on the information they receive through their senses to survive and function in an environment that is constantly changing. However, incoming information from the senses is often ambiguous, noisy, or incomplete. Context, either spatial, temporal, or cognitive, is fundamental, especially in visual perception, to disambiguate and complete this noisy input and optimize behavior. The mechanisms by which contextual information influences visual perception, however, are not fully understood. Studies in the current thesis aim to add to the understanding of those mechanisms. It comprises two lines of work. The first line of work focuses on the spatial context, where using fMRI, we investigate the neural correlates of spatiotemporal properties of context-dependent lightness perception. Results show that activity in the primary visual cortex (V1) correlates with context-dependent lightness perception, providing evidence for low-level mechanisms underlying the contextual effects. The second line of work focuses on the cognitive context, where we systematically study the effect of expectations about dynamic material properties on perceptual decisions. To do so, we used behavioral methods, where we manipulated participants' long-term and short-term expectations about material properties. Results show that expectations about material properties can impact relatively low-level perceptual decision-making processes. Furthermore, we found an interplay between long-term and newly learned expectations. In conclusion, the current thesis broadens our understanding of how context influences visual processes, particularly by providing evidence that low-level processes are affected by the visual context. This knowledge has the potential to help develop more accurate models of visual perception, which in turn can have implications in clinical neuroscience, artificial intelligence, computer vision, and marketing.

Keywords: Contextual effects, Lightness induction, Visual perception, Expectations, Perception of material properties.

ÖZET

BALAM VE BEKLENTĐLERĐN GÖRSEL ĐLEME DĐNAMĐKLERĐ ÜZERĐNDEKĐ ETKĐLERĐ

Amna Malik
Nörobilim, Doktora
Tez Danışmanı: Hüseyin Boyacı
Eylül 2023

Canlı organizmalar, sürekli değişen bir ortamda hayatta kalabilmek ve işlevlerini yerine getirebilmek için duyuları aracılığıyla aldıkları bilgilere güvenirler. Ancak duyulardan gelen bilgiler genellikle belirsiz, gürültülü veya eksiktir. Uzamsal, zamansal veya bilişsel bağlam, özellikle görsel algıda, bu gürültülü girdiyi netleştirmek, tamamlamak ve davranışı optimize etmek için temel bir unsurdur. Ancak bağlamsal bilginin görsel algıyı etkilediği mekanizmalar tam olarak anlaşılamamıştır. Mevcut tezdeki çalışmalar bu mekanizmaların anlaşılmasına katkıda bulunmayı amaçlamaktadır. Bu tez iki çalışmadan oluşur. Birincisi, fMRI kullanarak bağlama bağlı açıklık algısının uzamsal-zamansal özelliklerinin sinirsel bağıntılarını araştırdığımız mekansal bağlama odaklanır. Sonuçlar, birincil görsel korteksteki (V1) aktivitenin bağlama bağlı açıklık algısı ile ilişkili olduğunu ve bağlamsal etkilerin altında yatan düşük seviyeli mekanizmalar için kanıt sağladığını göstermektedir. İkinci çalışma, dinamik malzeme özelliklerine ilişkin beklentilerin algısal kararlar üzerindeki etkisini sistematik olarak incelediğimiz bilişsel bağlama odaklanır. Bunu yapmak için katılımcıların malzeme özelliklerine ilişkin uzun vadeli ve kısa vadeli beklentilerini manipüle ettiğimiz davranışsal yöntemler kullandık. Sonuçlar, malzeme özelliklerine ilişkin beklentilerin nispeten düşük seviyeli algısal karar verme süreçlerini etkileyebileceğini göstermektedir. Ayrıca, uzun vadeli beklentiler ile yeni öğrenilen beklentiler arasında bir etkileşim bulduk. Sonuç olarak mevcut tez, özellikle düşük seviyeli süreçlerin görsel bağlamdan etkilendiğine dair kanıtlar sağlayarak, bağlamın görsel süreçleri nasıl etkilediğine dair anlayışımızı genişletmektedir. Bu bilgi, görsel algının daha doğru modellerinin geliştirilmesine yardımcı olma potansiyeline sahiptir ve bu da klinik sinir bilimi, yapay zeka, bilgisayarlı görü ve pazarlama alanlarında sonuçlar doğurabilir.

Anahtar sözcükler: Bağlamsal etkiler, Açıklık indüksiyonu, Görsel algı, Beklentiler, Malzeme özelliklerinin algılanması.

Acknowledgement

First and foremost, I would like to express my special appreciation and thanks to my advisor Prof. Hüseyin Boyacı, for his unwavering support, constant encouragement, and exceptional mentorship. I found great inspiration in his work ethic and positive outlook on life. Whatever I am today as a researcher is because of the time and effort he has put into me. He has shown me, by his example, what a good scientist and a good person should be. I am at a loss for words to truly express how his mentorship and guidance have changed my life, but I hope I will make him proud in every step of my scientific career.

My sincere gratitude also goes to Prof. Katja Doerschner for her support and assistance not only during my short research visit to JLU Giessen but also extending it throughout my project. It is because of her that I was able to complete a whole study based on my short secondment.

I extend my sincere appreciation to Prof. Burcu Ayşen Ürgen and Prof. Didem Kadıhasanoğlu for being on my thesis committee and for their valuable comments and contributions over the years. I would also like to thank Prof. Zahide Pamir and Prof. Buse Merve Urgen for graciously accepting the responsibility of being on my thesis jury and for their valuable time and feedback. I thank all my jury members for their insightful suggestions during my PhD viva.

I would like to thank all my colleagues and lab mates who provided an enriching academic environment and engaging discussions. Their diverse perspectives greatly contributed to the development of this research. It was an invaluable opportunity to be a part of the Computational and Biological Vision Group (CBVG). I want to express a special word of thanks to my dear lab friends, Merve and Dilara, for their unwavering support, particularly during the past year. Their friendship and assistance have been a tremendous source of strength.

I would like to extend my best regards to my all friends who have been an integral part of my journey at Bilkent University. Special Thanks to Liala Tul Qadr and Khushbakht Ali Qamar, who accompanied me along this challenging journey, supported me in the worst of my times and celebrated me in the best of my times. I cannot overstate how crucial their support has been, especially during the trying times of the COVID-19 pandemic. Speaking of the COVID

era, I must acknowledge my HSL friends (Sarosh, Subhani, Ahmad bhai, Azeem Bhai, Saad, and Sonia) who were just a call away, despite being in different parts of the world. Our conversations weren't just a source of comfort but also played a big part in helping me grow as a person. I want to give a big shout-out to my friend Ahmad bhai, whose friendship feels like having a real brother by my side. Lastly, I can't forget to mention my buddies from my final years, Maryam, Hilal, and Talha. I owe Maryam a ton for sticking around during my absolute worst sleepless nights. Bilkent not only set me up for career growth but also blessed me with these amazing friendships that I'll treasure forever.

I would like to thank my family, especially my two amazing sisters, Sana and Atiba, for their continuous support and endless love. I want to express my heart-felt gratitude to my mom for her unwavering love, support, and encouragement throughout my life's journey. Her sacrifices, guidance, and boundless care have been the foundation upon which I've built my dreams and aspirations. Mom, your belief in me has been my greatest source of strength, and I am deeply thankful for your constant presence in my life. Your love is a guiding light that has shaped who I am today, and I am forever grateful for all that you've done for me.

From the first day of my Ph.D. year till my graduation, there was this one person I cannot possibly thank and love enough, my incredible husband, Zaeem. When there was no one, he was always there for me. He believed in me when I didn't believe in myself. I cannot imagine myself completing this journey without his emotional support. And it wasn't just about the emotional backing; his academic insights and guidance were invaluable. I am so lucky that such an extraordinary researcher happens to be my husband. His dedication and love for science have been a tremendous influence on me, and I attribute my passion for science to him. So, Zaeem, from the bottom of my heart, I owe you. You made this Ph.D. rollercoaster not just survivable but actually enjoyable. Here's to more adventures together, both in academia and in life!

Contents

1 Introduction **1**

1.1 Role of context in perception 2

1.2 Scope and motivation of the thesis 4

1.3 Overview of thesis 6

2 Role of spatial context in dynamics of visual processing **7**

2.1 Introduction 7

2.2 Processing of lightness in context 11

2.2.1 Historical review of theories in lightness perception 13

2.2.2 Spatiotemporal properties of lightness induction 19

2.2.3 Neural correlates of context-dependent lightness perception 22

2.3 Neural correlates of spatiotemporal properties of lightness induction 28

2.3.1 Introduction 28

2.4 Materials and Methods 32

2.4.1 Participants 32

2.4.2 Behavioral Experiment 32

2.4.3 fMRI Experiment 35

2.5 Results 40

2.5.1 Behavioral experiment 40

2.5.2 fMRI experiment 41

2.6 Discussion 47

2.6.1 Spatiotemporal properties of simultaneous lightness induction 47

2.6.2 Neural correlates of spatiotemporal properties of simultaneous lightness induction 48

2.6.3 The neural mechanism of context-dependent lightness perception . . . 51
2.6.4 Conclusion . . . 53
2.7 Future Directions . . . 54
2.7.1 An effective connectivity study . . . 54
2.7.2 Role of higher areas . . . 55
2.7.3 Genralizibility across different stimuli . . . 55

3 Effect of expectations on the perception of dynamic material properties 57
3.1 Introduction . . . 57
3.1.1 Importance of expectations in visual perception . . . 60
3.1.2 Long-term and short-term expectations . . . 66
3.1.3 Computational basis of expectations - Bayesian Model . . . 69
3.1.4 Neural basis of expectations . . . 73
3.2 Material perception and expectations . . . 75
3.2.1 Living in a material world . . . 75
3.2.2 How do we perceive Stuff? . . . 76
3.3 Unmet Expectations About Material Properties Delay Perceptual Decisions- Behavioral study. . . . 79
3.3.1 Materials and Methods . . . 82
3.3.2 Results . . . 87
3.3.3 Discussion . . . 92
3.3.4 Conclusion . . . 100
3.4 Future Directions . . . 104

4 General Discussion 106

List of Figures

1.1 A) Spatial context: A gradient in the outer rectangle makes the uniform inner rectangle perceived as having a gradient in the opposite direction. B) Cognitive context: We are able to interpret missing or incomplete information based on our prior life experiences. C) Temporal context: You will anticipate a rainy day following the presence of dark clouds in the sky using prior knowledge about the world where rain is always temporally preceded by dark clouds in the sky. 3

2.1 Examples showing that spatial context influences visual perception. A) Ebbinghaus illusion: The size of the central disk surrounded by a ring of bigger disks is perceived as smaller than the disk of the same radius surrounded by a ring of smaller disks. B) Delboeuf illusion: The size of the disk encircled by a circle of a bigger radius appears smaller than the disk of the same radius encircled by a circle of a smaller radius. C) Tilt illusions: Central vertical gratings appear tilted anticlockwise due to the clockwise orientation of the surround grating. D) The Zöllner illusion: Two vertical lines which are parallel to each other are not perceived as parallel due to the presence of V-shaped lines that cross them. E) Simultaneous lightness contrast: A region embedded in a darker surround appears lighter than the region of the same luminance embedded in a lighter surround. 10

2.2 Adelson's checkerboard illusion. The amount of light reflected from squares labeled A and B is equal, as can be seen on the left panel. Yet square A appears to be darker than square B due to the difference in the spatial context. 11
2.3 An illustration of White's illusion. The patch on the black bar appears lighter than the patch of the same luminance on the white bar. 12
2.4 An illustration of Benary's illusion. Two triangles have the same luminance but the one on the top left appears lighter than the one on the bottom right. 14
2.5 Lightness assimilation effect. The grey bars embedded in a darker surround appear darker than those of the same luminance embedded in a lighter surround. Unlike the lightness contrast effect, perceived lightness is shifted towards the surrounding region and observed with high spatial frequency stimulus. 16

2.6 First, in a behavioral experiment, we measured the magnitude of induction using the dynamic version of SLI. Based on the results in the literature, we expect that induction decreases with the temporal frequency of the surround modulation. Next, in an fMRI experiment, we measured the fMRI responses under three stimulus conditions and at different modulation frequencies. The 'surround modulation' condition is similar to the stimulus in the behavioral experiment, which leads to the dynamic SLI effect at low but not high frequencies. In the 'disk-modulation' condition, the surround remains constant, and the disk luminance is modulated. In the 'control' condition, the surround is modulated as in the surround-modulation condition, but the disk is black, leading to no dynamic SLI effect. The right panel shows the possible outcomes. If a cortical region is involved in context-dependent lightness processing, the fMRI responses under the surround-modulation condition at low frequencies would be larger than those at high frequencies. Further, the responses under the surround-modulation condition would be larger than those under the control condition at low frequencies. The first row shows this possible outcome. Any outcome inconsistent with this pattern would call into question the involvement of that area in context-dependent lightness processing. The next two rows show two such possible outcomes. The first one would be from a region whose activity is not correlated with lightness and, at best, directly responds to luminance modulations in the surround. The bottom row shows the outcome if the region responds to the border contrast variation, not to the lightness variations in the disk. 31

2.7 A trial in the behavioral experiment was comprised of an induction and a response phase. During the induction phase, we presented the dynamic SLI stimulus at different surround-modulation frequencies ranging from 0.25 to 16 Hz. Participants were instructed to fixate on the fixation mark and judge the perceived lightness variation in a target region marked by thin concentring circles. The right column shows the three target regions, border (8°-11°), middle (4°-7°), and center (0°-3°). 33

2.8 Functional localizer stimuli. High-contrast counter-phase flickering checkerboard patterns were presented in pairs of arcs on a uniform gray background (arcs are overlaid on the experimental stimulus in this figure to provide a visual reference to the readers). Arcs 1, 2, 3, and 4 were used to localize the center (1.5°- 2.5°), middle (5°-6.5°), border (9°- 10.5°) and surround (14°- 16°) ROIs in V1, V2, and V3. For V4, only border and center ROIs were extracted. . . 38

2.9 Behavioral results. Left panel: Group average of the dynamic simultaneous lightness induction (SLI) as a function of surround modulation frequency for each eccentricity condition. SLI is higher at the outer eccentricity, which is closest to the border between the disk and the surround. In the middle and center eccentricities, SLI is highest at lower surround modulation frequencies, decreases with frequency, and nearly disappears for the highest frequencies. Shaded regions show the standard error around the mean. Right Panel: Group average of the dynamic simultaneous lightness induction (SLI) as a function of eccentricity for each frequency condition. 41

2.10 V1 group average of fMRI responses as a function of frequency in all four ROIs. fMRI responses near the border between the disk and surround are the same for the surround-modulation and disk-modulation conditions for all frequencies (Border ROI plot). This shows that those responses are likely caused by the contrast modulation at the border. Whereas in the central and middle ROIs, fMRI responses under the surround-modulation and disk-modulation conditions remain similar only at 1, 2, and 4Hz, but diverge at 8Hz (Central ROI and Middle ROI plots), supporting that V1 responses in those ROIs are correlated with SLI, not simply the contrast modulation at the border. Error bars: SEM. 43

2.11 **A**. fMRI responses in the center ROI under surround-modulation and disk-modulation conditions, replotted from Figure 2.10 center ROI plot. * indicate a significant difference under the pairwise comparison of estimated marginal means. **B**. The difference between responses under the two conditions, surround-modulation minus disk-modulation, as a function of frequency. 44

2.12 fMRI responses in the V1 center ROI under the surround-modulation, disk-modulation, and control conditions. At low frequencies, responses under the surround-modulation condition are approximately the same as those under the disk-modulation condition and larger than those under the control condition. Whereas, at high frequencies, especially at 8Hz, this relation reverses. At 8 Hz, the responses under the surround-modulation condition are approximately the same as those under the control condition and lower than those under the disk-modulation condition. This is exactly what would be expected from an area whose activity is correlated with the behavioral lightness induction effect, as depicted in Figure 2.6. Error bars represent SEM. 45

2.13 fMRI responses in V2, V3, and V4 center ROIs. Because it was not possible to localize the center ROIs in all participants, the data were averaged across 8 participants in V2 and V3 and across 7 participants in V4. Responses under the disk-modulation and surround-modulation conditions did not differ in any area, and they were larger than those under the control condition. The pattern of results suggests that the fMRI response is dominated by contrast variations at the border, as depicted in Figure 2.6. Error bars represent SEM. 46

2.14 Causality study: There should be a causal relation from border ROI to central ROI if lightness is calculated indirectly from edges at the center of the disk (filling-in mechanism). However, if this is not the case, there should not be a causal relation between the border and central ROI. 55

3.1 An example demonstrating the role of expectations in recognizing objects when the object itself is ambiguous. The left column shows three objects, A, B, and C representing a could, a street light, and a tree which are unrecognizable. The image on the right shows the whole scene from which the left image parts are taken from the locations labeled A, B, and C. All objects are recognizable when put into the context, even though the whole image is blurred. . . 62

3.2 An example demonstrating the role of expectations in visual search in a complicated scene. The left column shows four objects (a cone, a car, a window, and a speed limit sign). The image on the right shows the whole scene from which the left image parts are taken. Finding the expected objects (a cone, a car, or a window) is easy. But the speed limit sign is at an unexpected location (look at the top left corner of the wall of a building). Therefore, it takes longer to find it. 63

3.3 An example demonstrating 'Light from above prior' as a long-term expectation. All shapes except the one in the middle of the bottom row are perceived as convex. This is due to our expectations from the real world where light (illumination source) generally illuminates from above. 67
3.4 An example demonstrating rabbit duck illusion as contextual (short-term expectations). The same ambiguous object is more likely to be perceived as a bunny at easter time and a duck during Thanksgiving time. 68
3.5 Bayesian framework for modeling the effect of expectations on perception. See text for details. 72
3.6 An example illustrating objects made of different materials (plastic and ceramic plaster) but appear similar in terms of their surface appearance. 77
3.7 An example illustrating the same material category 'textile' in four pictures but all are having different visual properties. Despite the differences in low-level visual properties, when observers are asked to categorize them, they will place them all under the 'textile' material category. 78
3.8 As soon as we see a teacup start falling down, our visual system predicts what might happen to it: if not caught, the cup will hit the ground and shatter. If the cup wrinkles as a piece of cloth upon hitting the ground, we will be surprised and even amazed. Because our expectations and the visual input mismatch. 80
3.9 Six objects used as stimuli and their natural and atypical behaviors. For the participants in the 'natural group,' natural behavior was predicted, and atypical behavior was surprising. For the participants in the 'atypical group,' atypical behavior was predicted, and natural behavior was surprising. These expectations were formed through a training session before the main experiment. See text and Figure 3.10 for details about the training procedure. . . 84

3.10 Experimental Design: Each participant underwent a training session before the experimental session. Participants in the natural group were trained on naturally behaving objects, while participants in the atypical group were trained on atypically behaving objects. In the experimental session, contextual expectations formed in training sessions were unmet in 20% of the trials. 85

3.11 Reaction times (RTs) from the training sessions of both groups as a function of trial number. The blue solid line shows mean RTs for the natural group (across 14 participants), and the pink line shows mean RTs for the atypical group (across 14 participants). Shaded regions are the 95 percent confidence interval around the mean. Thick horizontal solid lines show mean RTs on the first and last 5 trials. Overall, RTs get shorter as the session progresses. This training effect is stronger for the participants in the atypical group, who were trained on the atypical behavior. 88

3.12 Mean RTs of predicted and surprising conditions averaged across participants. When expectations are unmet, whether natural or atypical, perceptual decisions are delayed (post hoc pairwise comparison of estimated marginal means are significant, *: $p < 0.01$). This effect tends to be stronger in the natural group. The error bars represent SEM. 90

3.13 Mean RTs per participant in the experimental session for (A) natural and (B) atypical groups. Light and dark blue bars denote mean RTs in the predicted and surprising conditions, respectively. Error bars correspond to SEM. 91

3.14 Mean RTs in the experimental session for breaking and non-breaking objects. The left side shows the results of the natural group, while the right side shows the results of the atypical group. Posthoc pairwise comparison of estimated marginal means is marked significant with ** for $p < 0.001$ and * for $p < 0.05$. Error bars represent SEM. For all significant pairwise comparisons, see Table 3.2. 94

3.15 This figure shows the first step where long-term priors are updated through training. The left panel shows modulation of strong long-term priors for both groups and the right panel shows modulation of weak long-term priors for both groups. Here predicted stimuli include shattering and rigid objects, while surprising stimuli include graveling and wobbling objects. 101

3.16 This figure shows the second step that how an iterative model may be able to predict the empirical results. 4 possible scenarios based on the strength of priors are depicted (explained in the discussion section). Here, the predicted condition can be shattering and rigid objects for the natural group and can be graveling and wobbling objects for the atypical group. While surprising can be graveling and wobbling for the natural group and can be shattering and rigid for the atypical group. 102

List of Tables

3.1 Linear mixed model fixed effect estimates from the experimental session. Estimates and standard errors correspond to Box-Cox transformed RTs. (Significant effects are in bold.) 89

3.2 Follow-up contrasts for the three-way interaction of condition, group, and object that we found in the experiment. Significant comparisons are marked with bold p value text. PRE: predicted, SUR: surprising, NAT: natural group, ATP: atypical group, BRK: breaking object, NBR: non-breaking object. 93

Chapter 1

Introduction

The environment organisms live in is always changing, and living beings rely on the information they receive through their senses to survive and function in this environment. From doing the smallest tasks, such as interacting with daily life objects, (for example, picking up a teacup), to the crucial tasks, such as determining the survival of an organism (for example, escaping a predator), incoming sensory information plays a very important role. The functionality of an organism depends on how accurate and precise this information is. However, incoming sensory information is often ambiguous, noisy, or missing due to many factors [1]. Some of them are:

Environmental factors: Factors such as lighting conditions in visual information or background noise in auditory information make input less clear and more challenging to interpret.

Limitations of sensory organs: Our sensory receptors are not always able to detect all the available information giving rise to missing or incomplete information. For instance, the limited capacity of peripheral vision in humans leads to reduced detail and color sensitivity in the periphery of the visual field.

Limitations of nervous system: The information transmitted to the nervous system is often noisy, for example, biophysical variability in the firing rates of neurons.

1.1 Role of context in perception

Context, in general, refers to the set of circumstances that provide additional information in an event. In visual neuroscience, it is additional information about the visual target of interest. Contexts are often considered as being constantly present, diffused, and multisensory in nature. As shown in Figure 1.1, context can be spatial, which is related to the physical location of the stimulus, or it can be temporal, which is related to the changes in the environment over time, or it can be cognitive which refers to the prior knowledge, expectations, and beliefs individuals hold regarding a situation or stimulus [2]. All of them are closely connected to each other and play a critical role in remembering past events, understanding current situations/scenes, and predicting future outcomes.

Context is essential to disambiguate the noisy input and provide it with a meaningful interpretation. Context helps us to process complex input by filtering out unnecessary information. For instance, when driving, we encounter enormous visual input. We use our contextual knowledge of what a road looks like, what a particular street sign indicates, and how fast or slow the other vehicles are, to simplify the perception and focus only on the most important elements. Contextual information is particularly important in visual perception as we are continually bombarded with the complex visual information which is used to navigate and interact with our environment. The spatial context in visual processing plays an essential role in a wide range of features, including size, depth, distance, orientation, and brightness perception [3, 4, 5]. For instance, in Figure 1.1-A, the inner rectangle has a constant luminance. Still, a gradient in the surrounding outer rectangle makes us perceive a gradient in the inner rectangle. Cognitive context facilitates visual perception by helping us recognize objects, shapes, and material properties, interpret scenes and guide our interaction with the environment. For instance, in Figure 1.1-B, we can interpret what this street sign is, based on our prior life experiences. Temporal context also helps us to interpret the changes in visual scenes over time [2]. The time stamps are stored in our brains whenever we encounter events. These time stamps enable us to establish a chronological sequence of events and are then used to predict the time and events. Figure 1.1-C

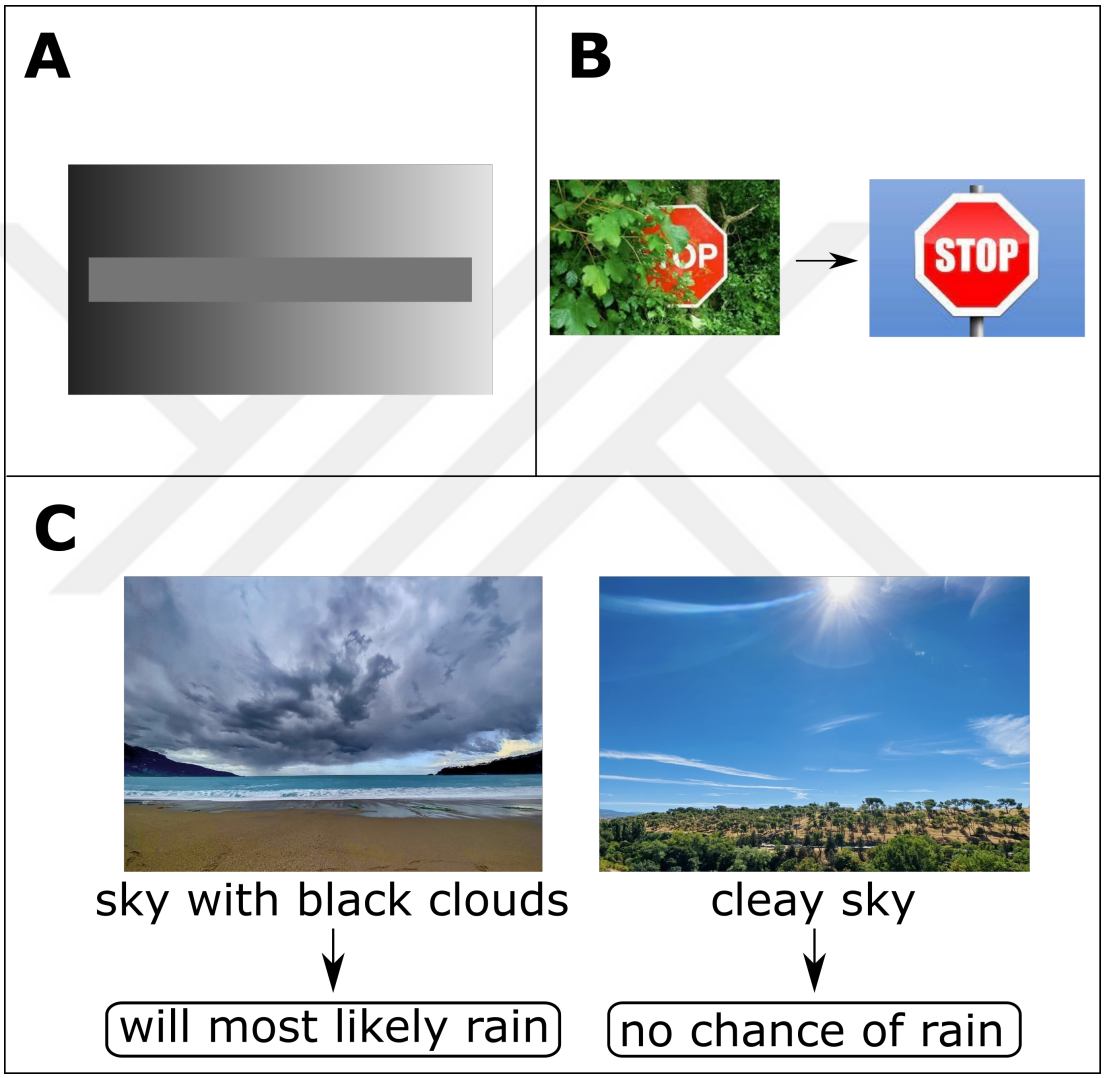

Figure 1.1: A) Spatial context: A gradient in the outer rectangle makes the uniform inner rectangle perceived as having a gradient in the opposite direction. B) Cognitive context: We are able to interpret missing or incomplete information based on our prior life experiences. C) Temporal context: You will anticipate a rainy day following the presence of dark clouds in the sky using prior knowledge about the world where rain is always temporally preceded by dark clouds in the sky.

shows how an observer can predict rain in the future by looking at the sky. Over time observer learns that the appearance of dark clouds is always followed by rain.

1.2 Scope and motivation of the thesis

Given the fundamental role contextual information plays in vision, a major scientific challenge is understanding how and where in the visual system these contextual effects come into play. Investigating contextual effects in visual processing can help elucidate the mechanisms by which our visual system integrates and modulates visual information based on surrounding context and prior experiences. It can help us to get deeper insights into the neural circuitry of the visual system and the role of feedback and feed-forward connections in general. Such knowledge will assist researchers in developing accurate models of visual processing. The inability to process contextual information can produce errors in perception and is also implicated in several psychopathological disorders [2]. Therefore, knowledge about the neural circuitry involved in contextual processing can be useful in developing interventions for these psychopathological disorders and in understanding perceptual illusions. Moreover, research on contextual effects in vision can have implications beyond neuroscience and psychology in computer vision and artificial intelligence. Mimicking human vision by taking into account contextual effects can help develop more realistic systems in robotics, computer vision, and autonomous vehicles. Thus, this thesis seeks to deepen our understanding of contextual effects in visual processing. This thesis focuses on the two types of contextual effects (spatial and cognitive) in separate lines of work.

The first project focused on the spatial context, where I investigated how contextual lightness information is processed in the visual hierarchy using a classical lightness illusion called simultaneous lightness induction (SLI). This illusion occurs when the perceived lightness of a region is affected by the context it is embedded in, as shown in Figure 2.1-E. Such illusions are used to study the neural

basis of lightness perception. However, despite enormous research, the mechanism and locus of processing of context-dependent lightness remain an important open question. Therefore, this study aimed to investigate the neural correlates of context-dependent lightness processing. For this purpose, I employed psychophysics and neuroimaging methods to investigate the effects of contextual processing of lightness in the primary visual cortex (V1), V2, V3, and V4. The current project adds to the literature by providing additional evidence for the role of V1 in contextual lightness processing while carefully accounting for the confounds that could be possible reasons behind contradictory results in previous studies.

The second project focused on the cognitive context, where I investigated the role of expectations in the perception of material properties in a dynamic scene. Perception of objects and their material properties are strongly affected by the cognitive context where our prior knowledge in the form of expectations is used to infer the properties of materials efficiently. There is a lot of literature on the effect of expectations on static stimuli, but a few studies have addressed the effects of expectations on material perception, especially in dynamic scenes. Hence, this project aimed to investigate how expectations about material properties can affect our decisions regarding the future states of objects. I employed a behavioral paradigm where I manipulated participants' long-term and short-term expectations about material properties. I investigated how it affected low-level perceptual decisions by examining participants' accuracy and reaction times. By showing that low-level perceptual decisions are delayed when expectations about material properties are unmet, this project broadens our understanding of the effect of expectations (cognitive context) on material properties and low-level perceptual decisions. This finding can potentially contribute to the development of computational models of perception. Moreover, this project can also have a broader impact in the field of marketing, where designers use such a surprise effect strategically in their products.

1.3 Overview of thesis

This thesis comprises four chapters. The first and current chapter gives a brief introduction and motivation behind the thesis, comprising two main projects.

Chapter 2 is focused on the first project, titled 'Role of spatial context in dynamics of visual processing', where I study the role of spatial context in lightness perception and its neural correlates. The chapter starts with an introduction of basic concepts and a literature review. I first present the methodology and results of the behavioral study and then proceed to the methodology and results of the fMRI study. Finally, I discuss our findings in light of the present literature.

Chapter 3 is focused on the second project, titled 'Effect of expectations on the perception of dynamic material properties', where I study the role of cognitive context on the perception of dynamic material properties. This chapter also starts with an introduction and a literature review. Then, I proceed to our experimental design and methodology. I present our results and discuss them in light of the given literature. Finally, in Chapter 4, I conclude the thesis with a general discussion.

Chapter 2

Role of spatial context in dynamics of visual processing

2.1 Introduction

The visual field is the area of space that is visible when the eyes are stably fixed in one direction [6]. The area of the visual field to which we direct our attention is called the visual target. Anything in the field other than the visual target provides visual context, mostly the area located right next to the target. This form of contextual information is called spatial context. As discussed in Chapter 1, context can also be temporal or cognitive. The focus of this chapter is spatial context, so I will be discussing contextual effects in light of spatial context.

Spatial context is ubiquitous in visual perception as objects or any visual target in the real world do not occur in isolation. When we look at any target, It will always have some spatial context, which may vary from time to time. How does our brain process the same target under different contexts? Will the target appear the same under a different context? Or will it be different? If it appears different, then what are the mechanisms governing this difference? What is the importance of spatial context in the dynamics of visual processing? These are some of the

questions I will discuss in light of the available literature in this chapter. I will start with some examples that illustrate the importance of spatial context and its effects on visual processing.

But before going into details of contextual effects, it is important to understand the distinction between the physical and perceptual attributes of an object/visual target. Perceptual attributes are the perceived physical attributes of an object/visual target. Physical attributes are objective measurements, for instance, the length of an object in centimeters, the size of a disk in degrees of visual angle, the orientation of a bar in degrees, or the distance between two objects in meters. They are measured by instruments and do not depend on the viewers. Perceptual attributes depend on the physical attributes either linearly (e.g., the perceived length of a bar increases linearly with the increase in actual length of the bar) or in a complex manner (e.g., perceived distance doesn't merely depend on the physical distance, but on many other visual cues such as retinal image size, texture gradient, binocular disparity, etc. [7]). The mapping between physical and perceptual attributes is not the point of interest here. But what is relevant under the current topic are the scenarios where the physical attributes of the target remain the same, but perceptual attributes change due to the change in the context in which the target is embedded. Intuitively thinking, whatever we perceive seems to correspond to the physical properties of the visual target we are looking at. But in reality, physical attributes of the context the target is embedded in also affect the perceived attributes of the target ranging from a weak to a critical effect [8]. Some examples of spatial contextual effects on low-level visual features are discussed in the section below.

Contextual effect on perceived size: Figure 2.1-A shows the Ebbinghaus illusion where the perceived size of the central target (disk) is affected by the ring of surrounding circles (inducer). The large size of the inducer makes the apparent size of the central disk smaller (left panel of Figure 2.1-A), and the small size of the inducer makes the apparent size of the central disk larger (right panel of Figure 2.1-A). The physical size of the central disk is same in both cases, but it is the spatial context that gives rise to the illusion [3]. Another similar illusion where context influences the perceived size is the Delboeuf illusion [9] as shown

in Figure 2.1-B. Many similar illusions, such as the Ponzo illusion, Mueller-Lyer illusion, and size-contrast illusion, demonstrate that objects' perceptual size is influenced by the context they are embedded in [10].

Contextual effect on perceived orientation: A typical example where context affects the perceived orientation is the tilt illusion as shown in Figure 2.1-C where the central grating appears to be tilted anti-clockwise away from the direction of the outer grating which is tilted clockwise when in fact the central grating is vertical [11]. In tilt illusion, the perceived orientation of the target is repelled away from the orientation of the surround grating when there is a small difference in angle between them [11, 4]. However, if the difference between them is large ($60° - 80°$), the target appears to be tilted toward the surrounding grating [12].

Another relevant example is the Zöllner illusion, in which lines in one orientation affect the perception of the orientation of lines in another direction, as shown in Figure 2.1-D. The two vertical lines in the figure are parallel, but the presence of a set of V-shaped lines gives this illusion of them not being parallel [13].

Contextual effect on perceived contrast: Several studies show that the background contrast does have an effect on the perceived contrast of the target [14, 15, 16, 17]. For instance, Chubb and colleagues have shown that if a textured pattern is embedded in a high contrast surround, the contrast of textures is perceived as lesser than it is [16]. This is called the Chubb illusion. Cannon and colleagues have demonstrated a similar effect for gratings [17].

Contextual effect on perceived lightness: The perceived lightness of a surface depends not only on its luminance but also on its context. For example, in the well-known simultaneous lightness induction (SLI) effect (Figure 2.1-E), a mid-grey central disk is perceived as lighter when embedded in a darker surround than when embedded in a lighter surround [5, 18]. SLI is an excellent illusion to study the contextual effects, as the magnitude of the illusion can be manipulated very easily by varying the properties of the surrounding region. Hence, the current study used a dynamic form of SLI to study contextual effects on lightness

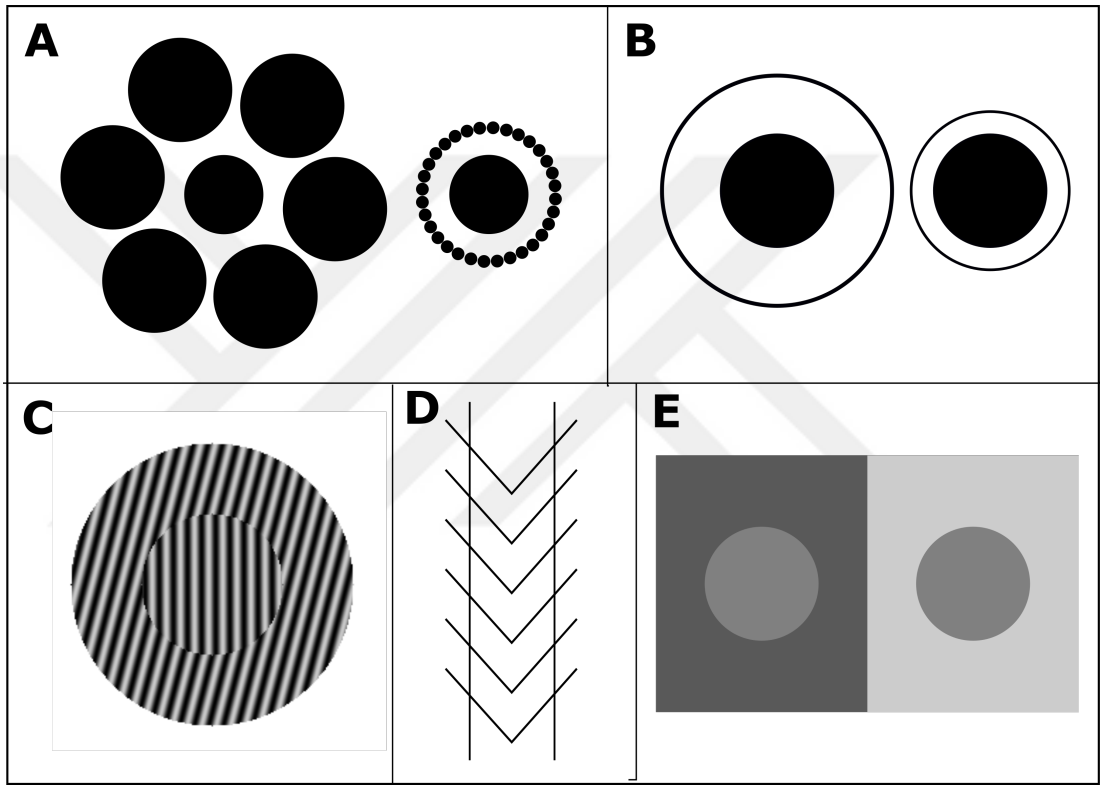

Figure 2.1: Examples showing that spatial context influences visual perception. A) Ebbinghaus illusion: The size of the central disk surrounded by a ring of bigger disks is perceived as smaller than the disk of the same radius surrounded by a ring of smaller disks. B) Delboeuf illusion: The size of the disk encircled by a circle of a bigger radius appears smaller than the disk of the same radius encircled by a circle of a smaller radius. C) Tilt illusions: Central vertical gratings appear tilted anticlockwise due to the clockwise orientation of the surround grating. D) The Zöllner illusion: Two vertical lines which are parallel to each other are not perceived as parallel due to the presence of V-shaped lines that cross them. E) Simultaneous lightness contrast: A region embedded in a darker surround appears lighter than the region of the same luminance embedded in a lighter surround.

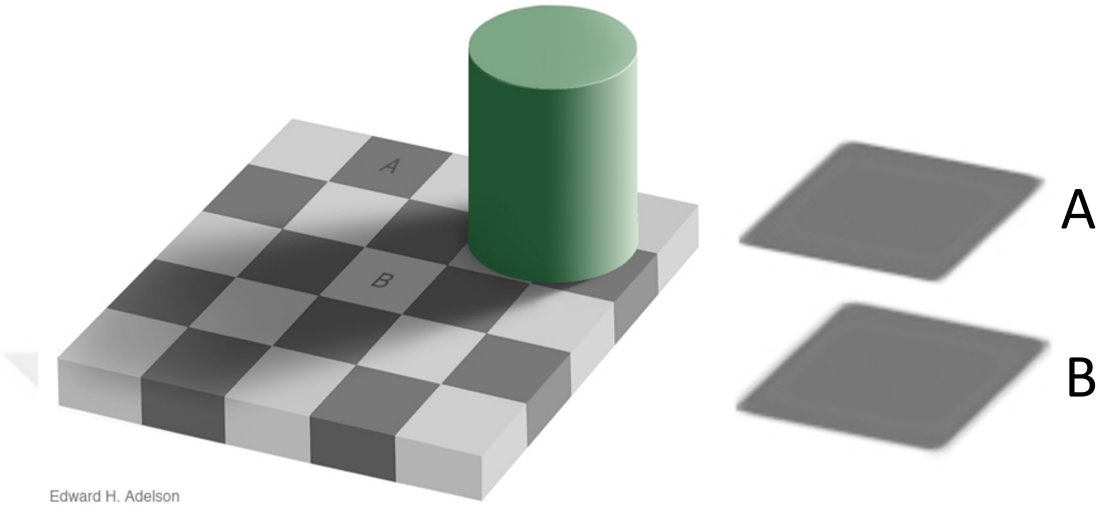

Figure 2.2: Adelson's checkerboard illusion. The amount of light reflected from squares labeled A and B is equal, as can be seen on the left panel. Yet square A appears to be darker than square B due to the difference in the spatial context.

processing, whereby varying the surround luminance at different frequencies, the magnitude of perceived lightness of the target/embedded region can be manipulated when in fact, the physical luminance of the region does not change. Since the current study focuses on lightness processing, we will discuss the contextual effects in lightness processing in detail in the section below.

2.2 Processing of lightness in context

In general, perceived lightness is the perceived achromatic shade of the surface, i.e., white, levels of gray, or black. This property holds enormous importance in neuroscience history and has been studied extensively because lightness is a property vital for object and material recognition. Its understanding may also elucidate the general mechanism brain uses to extract contextual influences in other domains. But before delving into details, I would highlight some key terminologies important in lightness literature. Surface reflectance refers to the ratio of light reflected by the surface to the light incident on the surface. Luminance is the amount of light emitted by a surface. Luminance varies as a function of

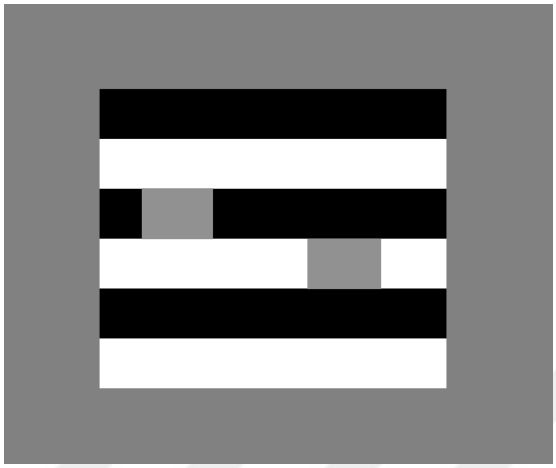

Figure 2.3: An illustration of White's illusion. The patch on the black bar appears lighter than the patch of the same luminance on the white bar.

surface reflectance and illumination. Lightness refers to the perceived relative reflectance, and brightness refers to the perceived luminance [19]. A surface can be perceived as white, gray, or black in terms of lightness and as bright, dim, or dark in terms of brightness. Reflectance and luminance are physical attributes of a surface, and lightness and brightness are its perceptual attributes. The reflectance of an object or surface is fixed, but luminance can vary depending on the illumination. Therefore, lightness is actually the relevant property for object recognition, and thus the visual system is more likely to primarily compute it. Hence, I will use the term lightness to discuss the contextual effects of surface reflectance perception in my study. However, terms can be used interchangeably when referring to literature in the following sections.

Two surfaces can have different luminance. However, the difference in luminance is not the point of interest; what is more crucial in the context of the current study is a scenario where two surfaces have the same luminance yet appear different in lightness because of the context they are embedded in, as shown in Figure 2.1-E (simultaneous lightness induction). Intuitively, we always assume that the lightness of a surface is directly related to the amount of light reflected from it. For instance, a white surface reflects more light than a black surface. However, lightness of a surface depends not only on the amount of light reflected

off the surface but also on the context it is embedded in. Some of the other common examples demonstrating contextual effects on lightness perception include Adelson's illusion [20] as shown in Figure 2.2. The amount of light reflected from squares labeled A and B is equal (see the left panel of Figure 2.2), yet square A appears to be darker than square B due to the difference in the spatial context. Another example is White's illusion [21] shown in Figure 2.3 where a grey patch on the black bar appears lighter than the lower grey patch of the same luminance on the white bar. Unlike SLI shown in Figure 2.1-E, both patches here share an equal amount of border contrast; hence it cannot be explained by differences in border contrast. Benary's illusion, as shown in Figure 2.4, is another common demonstration where the small triangles have identical luminance but still appear different in lightness due to contextual effects [22].

Hence, all these examples demonstrate that context has a profound influence on the processing of surface lightness or brightness. Such context-dependent effects are frequently used in the literature to study lightness perception and its neural correlates because they allow researchers to dissociate lightness and luminance and help to understand mechanisms behind the processing of contextual information in general [23, 24, 25, 26, 27, 28, 29, 30, 31].

2.2.1 Historical review of theories in lightness perception

Work on lightness perception holds a very long history [32, 33, 34, 35, 36]. One of the earliest descriptions of simultaneous contrast phenomenon dates back to the eleventh century by Alhazen (1025) as follows:

> *"Thus if a pure white body is dotted with a dark-coloured paint by allowing small drops of the paint to fall on it, or if minute designs are made on it with this paint, the paint will look black or very dark; its distinctive quality will cease to be apparent, and the eye will not be able to perceive its true colour. If marks are made with the same paint on a pitch-black body, the paint will look white or pale-coloured;*

Figure 2.4: An illustration of Benary's illusion. Two triangles have the same luminance but the one on the top left appears lighter than the one on the bottom right.

> *its darkness will not be apparent, and the eye will fail to perceive its true colour. If however, this paint is placed in midst of bodies that are not extremely white or extremely black, its colour will appear as it is and the eye will perceive its true colour in so far as it can be perceived by sight .... For the qualities of lights and colours are perceived by the eye only by comparing them with one another."*

Another notable historical figure from the sixteenth century was Leonardo da Vinci [32], an artist with a keen understanding of light and shadows. Within his paintings, da Vinci skillfully manipulated the interplay between luminance and size, crafting an illusion of depth and realism [32], leading to a phenomenon that was systematically studied later by vision scientists [33]. With subsequent contributions of numerous scientists across centuries, research seeking to discover the brain areas and mechanisms involved in the neural processing of lightness continues to this day. However, the results of these studies have not yet led to a consensus. Below, I summarize the literature on low, mid, and high-level accounts of lightness perception.

2.2.1.1 Low-level account:

Lateral inhibition and spatial filtering: The concept of the receptive field was introduced by Hartline [37] in 1940, which supported a feed-forward neuronal model for visual processing. Inspired by this work, Hubel and Wiesel [38] proposed the excitatory-inhibitory structure of ganglion cell receptive field. Following up by extensive research on the center-surround receptive field, Enroth and Robson [39] proposed a mathematical model of the receptive field called the difference of Gaussian (DOG) model comprising two separate Gaussian functions for the center and surround. This model was successful in explaining lightness contrast effects by lateral inhibition. According to this model, the ganglion cells in the retina are not activated by the luminance per se but by the spatial change in the luminance between the center-surround receptive field. Hence, a central region embedded in the lighter surround produces less contrast than the same region embedded in the darker surround (Figure 2.1-E), giving rise to small responses in neurons [40]. This model sufficiently explained contrast effects but was not able to explain another common form of the contextual effect on perceived lightness called the lightness assimilation effect (Figure 2.5) as well as other lightness illusions, e.g., White's illusion, grating induction, and Wertheimer-Benary effect. Moreover, numerous studies demonstrated that the lightness contrast phenomenon could be observed with the test regions larger than the classical receptive fields of the retina and LGN, which implied that a simple DOG model was not sufficient to explain contextual effects in lightness processing [41, 42, 43, 44]. Later, many researchers attempted to refine the DOG model to cater to these limitations [45, 46, 47], which led Blakeslee and McCourt [27] to develop a multiscale spatial filtering model in the primary visual cortex which was constructed of multiple DOG filters at different scales of spatial frequency. This model explained the lightness contrast effect (observed with low spatial frequency), lightness assimilation effect (observed with high spatial frequency), and grating induction. However, this model still couldn't account for White's illusion. Later, Blakeslee and colleagues [48, 49, 50] refined the model further by adding a component for orientation as well as global contrast normalization to explain White's illusion.

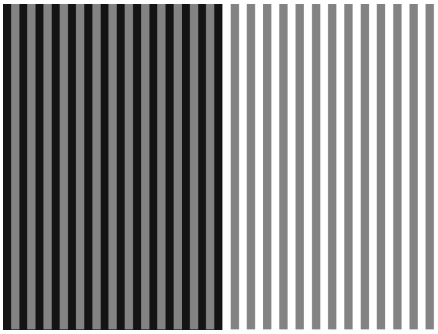

Figure 2.5: Lightness assimilation effect. The grey bars embedded in a darker surround appear darker than those of the same luminance embedded in a lighter surround. Unlike the lightness contrast effect, perceived lightness is shifted towards the surrounding region and observed with high spatial frequency stimulus.

A major challenge to multi-scale spatial filtering account to date is that it fails to explain lightness constancy in complex scenes with heterogeneous illumination. It cannot differentiate between the luminance changes caused by shading from those caused by changes in the surface reflectance. Hence, It can only predict the perceived brightness but does not attempt to predict the surface reflectance. Secondly, an asymmetry of lightness induction has been observed across increments and decrements i.e. darker surrounds produce a bigger lightness induction effect than the lighter surround [48]. Multi-scale spatial filtering account cannot predict this asymmetry [51].

Filling-in mechanism: Another competing low-level account used to explain lightness processing is the filling-in mechanism. According to this theory, it is the lightness information at the border that is detected by the neurons and then propagated to the neurons corresponding to the region to be filled in [52]. From the studies of pathological scotomata and blindspot, it has been known for a long time that perceptual filling-in plays an essential role in completing the missing information in the visual field [52, 53]. Artificial scotomata can be produced by stabilizing the image. A while after watching a stabilized image, its brightness and color fade, creating a blind spot. Gerrits and colleagues [54] created artificial scotomata and found that the surrounding color fills in the blind spot. Many other similar studies showed that it is the information at the edges that plays a vital role in determining the brightness or color of the bound region [55, 56, 57, 58].

The primary source of evidence supporting this theory comes from studies that investigate temporal properties of lightness processing and show that it is a slow process [23, 58, 59, 60, 61]. The details are discussed in following subsection where we talk about spatiotemporal properties of lightness induction. However, the findings in the literature on temporal properties of lightness induction are inconsistent. Some studies [62] show that lightness perception is instantaneous, challenging the filling-in account. A major challenge to filling-in account is explaining the spatial nonuniformity in the perceived lightness across the enclosed surface, which will also be discussed later in detail [62].

2.2.1.2 Mid level to higher level account:

Higher level account for vision dates back to the 1700s to the work of George Bishop Berkeley [63] who proposed the idea of immaterialism in his book called *'An Essay towards a New Theory of Vision'*. He argued that retinal images themselves carry insufficient spatial information and require additional interpretation. In the dialogue *'Three Dialogues between Hylas and Philonous'* published later in 1710 to address his critics [64], Berklay states:

> *"Light and colors, as immediately perceived by us, I grant cannot exist without the mind. But in themselves, they are only the motions and configurations of certain insensible particles of matter."*

Unconscious inference to infer lightness: It was a century and a half later in 1866 when Helmholtz [33] defined vision as unconscious inference in his *'Handbook on Physiological Optics'*. His work was one of the earliest attempts to solve the problem of lightness constancy. He suggested a higher-level mechanism in the form of unconscious inference is used to estimate the reflectance of a surface from the mean illumination of a scene. However, Hering [35] criticized the high-level account of Helmholtz, arguing that low-level mechanisms are sufficient to explain lightness perception emphasizing the role of lateral inhibition, pupil size, and adaptation. Later, David Katz [65] opposed both Helmholtz and Hering showing

that lightness constancy is observed even if all of Hering's suggested mechanisms, as well as color memory, are ruled out.

Intrinsic image theories: Meanwhile, Gestalt psychologists [66, 34, 67, 68, 69, 70, 71] came up with a mid-level account, rejecting both low-level accounts and high-level accounts. Such accounts emphasized the importance of gradients/relative luminance, grouping, and depth perception in estimating lightness in complex scenes. These are called intrinsic image theories. Some noteworthy studies pertaining to intrinsic image theory were by Gelb [34], who showed that a piece of black paper appears much darker in the spotlight when presented along with real white than when placed alone in the spotlight, Koffka [67], who was the first to propose a theory of lightness based on luminance ratios, Wallach [66], who showed that two disks of different luminance are perceived to be the same when they are presented with equal disk-to-background luminance ratios. Wallach [66] emphasized the role of luminance ratios but not absolute luminance in perception. The luminance ratio rule of Wallach gained a lot of attention. But it accounted only for illumination-independent constancy and failed to explain another type of lightness constancy called background-independent constancy, which led to the development of edge integration theories suggesting that the visual system is capable of estimating the relative lightness of surfaces over large distances [72, 73, 74]. Gilchrist [36] later refined edge integration theory suggesting that such edge integration occurs after the classification of edges into either reflectance edges or illuminance edges.

Anchoring theory: One of the limitations of edge integration theories was the lack of an anchor in inferring the luminance from relative luminance ratios. Gilchrist [75] proposed his own theory called anchoring theory to address this limitation. According to his model, an image consists of local and global frameworks; the highest luminance in a framework acts as an anchor. The reflectance of other regions is estimated by comparing it to this anchor. Finally, the perceived reflectance of a region is estimated by weighted averaging of local and global reflectances [75]. The main drawback of anchoring theory is the lack of a computational model [76].

Symbolic theory: Another notable theory supporting high-level account is symbolic theory [77, 78]. This theory is based on the filling-in mechanism hence also called the symbolic filling-in mechanism. According to this theory, it is the signal at the edges of an image that determines its surface lightness but this signal is first sent to higher-level areas, which then process this signal to estimate lightness [78]. Perna and colleagues [77] conducted an fMRI study where they found evidence for filling-in but did not find significant activity in V1 correlated with filling-in. Instead, they demonstrated that activity in higher visual areas correlated with lightness, supporting the symbolic filling-in theory.

Aside from the limitations and strengths of all the theories discussed above, none of them could explain how and where our visual system implements these computations.

2.2.2 Spatiotemporal properties of lightness induction

Investigation of spatiotemporal properties of lightness induction can potentially give essential insights into the mechanisms and neural underpinnings of lightness processing and hence have been investigated by several studies [49, 23, 29, 79, 80, 61, 62, 30].

Spatial properties: One of the consistent findings from all the studies on spatial properties of perceived lightness is that as the distance from the inducing field is increased, there is a decrease in the magnitude of lightness induction [49, 50, 80, 79]. In the quest to find the effects of the configuration of stimulus, speed, and magnitude of luminance change, and stimulus size and shape on lightness induction, Paradiso and Hahn [61] conducted a study where they presented a spot of light having constant luminance but changed the luminance of the background. They found that changing the luminance of the background produces a change in the perceived lightness of the spot, but this change is not uniform within the spot, and lightness seems to sweep inwards from the boundary of the spot. Paradiso and Hahn [61] suggested that the slow process like filling-in can be a potential explanation behind this lag and its dependence on the rate

of luminance change of surround. Pereverzeva and Murray [29] investigated the neural correlates of lightness induction using fMRI using a disk of constant luminance embedded in a surround that was temporally modulated at 1 Hz. They found that the magnitude of the lightness induction effect reported by observers is greater at the border and decreases toward the center. This decrease was also reflected in the BOLD signal in V1. Similar inhomogeneities in perceived lightness have also been reported in studies using static stimuli [81, 82]. Later, Blakeslee and McCourt [83] used a novel quadrature phase motion cancellation technique to measure the strength of induction as a function of distance from the inducing field edge and confirmed that magnitude of induction decreases as the distance from the inducing edge increases. However, there is evidence that flicker sensitivity varies across the retina from the periphery to the fovea [84, 85], which raised a possibility that the observed inhomogeneity in induction magnitude could be the result of the difference of flicker sensitivity across different locations. Pereverzeva and Bromfield [79] conducted another behavioral study using a stimulus similar to the one used by Pereverzeva and Murray [29] to rule out the possibility that any spatial non-uniformity observed in perceived lightness was due to different retinal positions. In one of the three experiments, they asked observers to directly look at the test patch (either at the border, middle, or center of the disk) and keep the retinal position the same but vary the distance from the inducing field. They confirmed that such spatial non-uniformities are still observed if the retinal position is fixed. These spatial inhomogeneities have been reported in grating induction and White's illusion, as well [27, 48]. What causes this spatial inhomogeneity is up for speculation. Multiscale spatial filtering models account for observed inhomogeneities [83]. This property has also been used to argue against a mechanism like filling-in because a filling-in mechanism predicts that the region which is filled in carries the same edge signal. Therefore, it produces a percept of a homogenous surface.

Temporal properties: The time course of lightness perception can give insights into the potential mechanisms involved. Therefore, temporal properties of lightness illusions have been thoroughly investigated. One of the first studies on temporal properties by Paradiso and Nakayama [58] demonstrated that backward

masking of a white disk on a darker surround with small white ring results in a darker perception of the center of the disk. They found that as the distance between the edge of the disk and the masking ring increased, the longest time delay at which the mask was effective increased. They implicated this percept as a result of the spreading of edge information and argued that as the distance is increased, the spread of edge signal takes more time. So, a temporal delay in the perception of lightness would be indicative of a filling-in mechanism. Using the lightness induction illusions, many other studies demonstrated a temporal delay in lightness perception [23, 86, 87]. On the other hand, Blakeslee and McCourt [62] argued against the filling-in mechanism in the study, where they show that lightness induction is an instantaneous process.

Another approach used in literature to study the temporal properties of lightness induction is the modulation of the luminance of the surrounding region while keeping the central region's luminance constant. Even though the luminance of the central region is constant, it is perceived to be flickering out of phase to the surrounding region. Studies using this approach show that as the frequency of the surround's luminance modulation increases, the magnitude of illusory flicker decreases [23, 86]. These studies support the filling-in process by arguing that at frequencies where the propagation speed of the filling-in signal is lower than the rate of change of surround modulation, the magnitude of induced lightness will decrease in the center. They also imply that the magnitude of induced lightness depends not only on the rate of change of surround luminance but also on the distance the signal has to travel. That is why the induced lightness in the center of the region decreases as the size of the induced region is increased [79]. The cut-off frequency for lightness induction is reported to be 2-3 Hz [23, 79], which predicts filling-in to be a very slow process with a velocity of $4^{\circ}/s$. However, the data on the temporal delay of lightness perception shows the process to be faster or even instantaneous [58, 62, 79, 88]. Thus, the proposed speed of filling-in mechanism is inconsistent in the literature.

2.2.3 Neural correlates of context-dependent lightness perception

2.2.3.1 Role of primary visual cortex

Early work on the neuronal receptive field responses from the primary visual cortex (V1) in cats and monkeys demonstrated that cells in V1 respond to a restricted area in the visual field to a particular stimulus, for example, a bar oriented in a particular direction [89, 90]. This area is called the classical receptive field. For the first time, Lamme [91] demonstrated that not only the stimuli inside the classical receptive field but those present outside may also modulate the activity of neurons in V1. He conducted an experiment on a figure-ground segregation task. The activity of V1 cells was recorded in awake monkeys while they saw a textured stimulus. He showed textures in such a way that the texture was always the same within the RF of the V1 cell being recorded, but the direction or motion of the texture presented outside RF varied from trial to trial. He found that the activity of cells is significantly modulated when the direction or motion of texture changes outside the classical RF. This was a dramatic demonstration of the extra classical receptive field effect, which led the researcher to find contextual influences in V1 in various other psychophysical processes, including contour integration [92], focal orientation discontinuities [93], speed perception [94], perceptual popout effect [95] and lightness perception [88]. Recall that a major criticism of low-level contrast mechanisms in lightness perception raised at this time was that the phenomenon of simultaneous lightness contrast could be observed in test regions that are larger than the sizes of classical receptive fields in V1, LGN, and retina [41, 42, 43, 44]. So, the discovery of extra-classical receptive field effects in V1 diverted the attention of researchers back to the low-level accounts. The underlying neural mechanism behind these contextual influences in V1 was suggested to be either lateral interactions within V1 or feedback from higher visual areas [96], and these influences were shown to be either facilitator, inhibitory or a mix of both [97, 98, 99, 100, 101, 102, 103, 104, 105]. Below, we briefly review some of this literature.

Psychophysics studies

Early psychophysics studies supporting low-level account of lightness induction suggested underlying mechanisms at the retinal level [21]. However, McCourt conducted a study in 1982 in which he demonstrated the phenomenon of grating induction [106]. The study showed that induced gratings are produced in a phase opposite to the inducing grating, and the magnitude of induced gratings decreased with an increase in spatial frequency. The spatial frequency and directional sensitivity provided evidence against the retinal involvement and in favor of cortical mechanism beyond the level of spatial frequency selective channels. Paradiso and Hahn [61] conducted an experiment in 1996 where they presented a luminance-modulating disk. They positioned the disk strategically to surround the blind spot of observers, extending an angle of $1.5° - 2.5°$ around it. They found that perceived lightness sweeps to the inside of the disk even though the central part of the disk is projected onto the part of the retina devoid of photoreceptors. Activated photoreceptors correspond only to the extended annular portion of the disk beyond the blind spot. This finding suggested that the neural correlates of perceived lightness exist beyond the retinal photoreceptors [61].

Electrophysiological studies

Several electrophysiological studies demonstrated the role of the primary visual cortex (V1) in processing surface luminance [107, 108, 109]. However, it was the work of Rossi and colleagues [24, 88] that first elucidated the role of the primary visual cortex (V1) in context-dependant processing of surface luminance. Using the dynamic version of the simultaneous lightness induction stimulus, Rossi and colleagues [24] investigated the neural correlates of the context dependant lightness processing at different temporal frequencies of surround luminance modulation. They found that temporal modulation of surround luminance modulated the activity of neurons in V1 in a way that correlated with perception. The cut-off frequency of surround modulation for the perception of lightness induction reported earlier was 4 Hz [23, 86]. Interestingly, in line with these results, Rossi and colleagues [24] did not find modulation of V1 activity when the frequency was

increased beyond 4 Hz. They found both types of neurons whose activity modulated out of phase with surround modulation and those whose activity modulated in phase with surround modulation. Kinoshita and Komatsu [110] conducted a similar study using a static version of stimuli used by Paradiso and colleagues [58] and found similar results.

Neuroimaging studies - Evidence against the role of V1

Some of the earliest fMRI studies investigating neural correlates of lightness perception argued against the role of V1 in perceived lightness [30, 77, 111]. One of these studies is by Perna et al. [77], who investigated the BOLD signal associated with Craik–Cornsweet–O'Brian illusion to elucidate brain areas involved in contextual lightness processing. They reported a strong response in the V1 corresponding to the retinotopic location of edges. They argued that this edge response was unrelated to illusory lightness as the same response was found for a line of matched contrast. Furthermore, no response to edge versus line stimulus was observed in the V1 corresponding to the retinotopic location of the center of the square. However, they did find a correlation between perceived lightness with higher areas. Another study by Boucard et al. [111] investigated BOLD signal in the retinotopically organized regions of the visual cortex (they did not differentiate the primary visual cortex from secondary areas) corresponding to a dynamic stimulus similar to the one used by Paradiso et al. [61] in their electrophysiological study of cats and reported a similar increase in activation for both to surround luminance modulation as well as disk luminance modulation. However, they found a large delay (approximately 3s) associated with surround modulation relative to the onset of lightness percept. In light of this finding, they argued against the V1 signal being associated with lightness induction. Using the same paradigm as Boucard et al., [111], Cornelissen et al. [30] have found no evidence of context-dependent lightness processing in V1. In that study, authors recorded V1 responses to a foveally presented static disk embedded in a surround whose luminance varied at a relatively low frequency (1 Hz). They found that, in line with the SLI effect, the perceived lightness of the disk varied as a function of time. When they analyzed the fMRI BOLD responses in regions of V1 that retinotopically process the part of the visual field occupied by the disk, however,

they found that the activity could be explained simply by long-range responses to the contrast variation at the border between the disk and the surround. Thus, they argued that there is no evidence of fMRI response to context-dependent lightness variations in V1 [30].

Neuroimaging studies - Evidence in favor of the role of V1

Haynes et al. [26] demonstrated that the BOLD signal in V1 correlates with the perception of surface luminance. In a psychophysical experiment, they measured the magnitude of perceived lightness and showed that it correlates with the fMRI responses. However, this study did not disentangle the luminance and lightness. Moreover, they didn't address anything about context-dependent lightness processing. Sasaki and Watanabe [112] showed evidence for the presence of color filling-in the BOLD signal in V1. Boyaci and colleagues [113] presented the first fMRI study to systematically report the role of the primary visual cortex in context-dependent lightness. They used Craik–Cornsweet–O'Brian stimulus and used three conditions; an illusory condition producing an illusory difference in flanks, a real condition in which flanks differed in physical luminance, and a control condition in which there was a very weak lightness effect. A positive BOLD signal was reported for both real and illusory conditions, implying that V1 processes both physical and illusory changes in luminance. By comparing the illusory condition with that of control, they ruled out the possibility that response in the illusory condition could be associated with long-range edge responses proposed by Cornelissen et al. [30]. This was followed up by another fMRI study by Pereverzeva and colleagues [29], who examined the V1 activity using the dynamic SLI and a clever design that allowed dissociating responses to lightness and border contrast variations. Specifically, they presented the center disk at different luminances, ranging from very low to mid values. Perceptually the SLI effect increased as the disk luminance increased, whereas temporally integrated border contrast energy decreased. Results showed that in regions of V1 that retinotopically correspond to the disk's center, there was a larger fMRI activity on trials where the central disk had higher luminance; in other words, the activity correlated with SLI, not border contrast. Thus, the study provided evidence that human V1 is involved in processing context-dependent lightness

[29].

One of the crucial considerations is that finding activity in early visual areas corresponding to perceived lightness by itself does not rule out the influence of higher visual areas in lightness perception via top-down influence as mentioned by Boayci et al. [113].

2.2.3.2 Role of higher visual cortex

In addition to V1, as discussed earlier, Boyaci et al. [113] found activity in V2 and V3 to be correlated with perceived lightness. Evidence for the V2 responses in perceived lightness comes from some other electrophysiological studies on monkeys and cats as well [114, 115]. Roe et al. [115] in an electrophysiological study in anesthetized monkeys found that activity in V2 correlates with perceived lightness as well as a physical change in luminance. However, unlike Boyaci et al. [113] they did not find V1 activations correlated with perceived lightness. Instead, they found V1 activity to be correlated only with the actual change in physical luminance. Hung et al. [114], using Craik–O'Brien–Cornsweet illusion, also demonstrated that cells in areas 17 and 18 in a cat's visual cortex respond to both actual changes in luminance and perceived changes in luminance. However, they showed that cells in area 18 (which corresponds to V2 in humans) predominantly respond to a perceived change in luminance, while cells in area 17 (which corresponds to V1 in humans) predominantly respond to an actual change in luminance.

Techniques using high temporal resolution might give better insights into the dynamics of neuronal activity related to perceived lightness in the visual cortex. Using high-resolution magnetoencephalography (MEG) on human observers Zhou and colleagues [116] studied a temporal contextual phenomenon of lightness perception called temporal context effect (TCE) in which if a short flash of light is presented in the presence of a long flash of light, the perceived lightness of short flash increases monotonically with onset asynchrony [117]. Zhou et al.

[116] found that neuronal activity in lateral occipital complex (LOC) and V4 correlated with perceived lightness peaking at 200ms to 250ms, and activity in early visual areas V1, V2, V4, and VO correlates with physical luminance peaking earlier at around 100ms. Based on this finding, they proposed that lightness coding involves both bottom-up and top-down processes, with the possibility of V4 playing an important role in integrating both processes to code perceived lightness. The importance of V4 in lightness coding demonstrated by authors here is in line with some other recent studies [118, 119, 77]. Lateral occipital complex (LOC) is also reported to be activated in an fMRI study of Craik–Cornsweet–O'Brian stimulus by Pernal et al. [77] in addition to other areas of dorsal pathway called intraparietal sulcus (IP).

The different results from different studies might be due to the different stimuli used, different experimental designs, the use of awake versus anesthetized subjects, differences in the species of subjects used, or the different methods of measuring neural activity (electrophysiological versus neuroimaging). Hence, in conclusion, the locus of processing lightness information remains a debate. Whether V1 represents perceived lightness or not, the role of higher visual areas in lightness perception are questions yet to be answered.

2.3 Neural correlates of spatiotemporal properties of lightness induction

The work presented below is published as a preprint: Malik A., Boyaci H (2023) Neural correlates of spatiotemporal properties of lightness induction [120].

2.3.1 Introduction

As discussed in detail above, lightness is defined as the perceived relative reflectance of a surface, and it depends not only on the amount of light reflected off the surface but also on its context. For example, in the well-known simultaneous lightness induction (SLI) effect (Figure 2.1-E), also known as simultaneous brightness contrast), a mid-grey patch is perceived as lighter when embedded in a darker surround than when embedded in a lighter surround [5, 121, 122, 123]. Such context-dependent effects are frequently used in the literature to study lightness perception and its neural correlates because they allow researchers to dissociate lightness and luminance [23, 24, 25, 26, 27, 28, 29, 30, 31]. Results of those and other studies in the literature, however, have not yet led to a consensus about the visual areas and neural mechanisms involved in the processing of lightness. The proposed mechanisms can be broadly grouped into three categories; a low-level account [48, 124, 125], a mid-level account [75, 126, 127, 128] and a high-level account [31, 77]. Low-level accounts argue that early retinotopic visual areas play an instrumental role in the process. In several studies, neural activity in those areas was indeed found to correlate with context-dependent lightness, providing evidence in favor of low-level accounts [113, 29, 88, 112]. Some other studies, however, provided opposite results [30, 77]. Thus, whether and how low-level visual areas, including the primary visual cortex (V1), contribute to lightness processing remains an important open question.

To answer this question, a dynamic version of the SLI effect, where a constant-luminance disk is perceived to vary in lightness when presented in a surround that

is temporally modulated in luminance, is frequently used. However, the results of those studies regarding the role of early retinotopic areas in perceived lightness are inconsistent, as described in the previous section [24, 29, 30]. One of the main reasons for these inconsistent results is that dissociating neural responses to perceived lightness from the responses to contrast is very challenging using fMRI. Pereverzeva and colleagues [29] tried to achieve this by using a clever design, as discussed previously. In their design, however, the luminance of the central disk varied between different conditions. This could have confounded the results because the increase in the perceptual lightness effect was accompanied by an increase in the mean luminance, and an increase in mean luminance, independent of contrast, may increase the neural responses in early visual areas [129, 130].

Here, we followed a similar strategy to dissociate responses to context-dependent lightness and border contrast without changing the luminance of the central disk. To do so, we kept the central disk luminance constant and used different surround modulation frequencies ranging from low to high values, and collected fMRI BOLD responses ('surround-modulation' condition). The dynamic SLI effect decreases with frequency [23, 79, 86], while border energy computed by integrating the contrast variation over time increases. Thus, we reasoned that this design should allow us to dissociate the responses to context-dependent lightness and border contrast and address the confound put forward by Cornelissen et al. [30] and others. Further, in order to compare responses to luminance-dependent and context-dependent lightness variations, we have introduced a 'disk-modulation' condition in which the surround remained static, and the disk varied in luminance with the same temporal modulation characteristics as in the surround-modulation condition. Because the temporal contrast energy at the border for a particular frequency was equal across the two conditions, the difference between disk-modulation and surround-modulation conditions should reflect mainly the difference between neural responses to luminance and context-dependent lightness variations.

We noted that surround modulation alone might drive the fMRI responses without any lightness-related neural activity. This could happen because the fMRI response has a wide point-spread function and reflects the activity of a large

number of functionally heterogeneous neurons in its spatial unit, i.e., a voxel. Even in a functionally identified cortical region of interest that retinotopically corresponds to the inside of the disk, there might be neural responses to the surround modulation or the overall illumination variation in the environment. To address this confound, we used a separate 'control' condition in which surround luminance varied as in the surround-modulation condition, but the disk was black, causing no lightness induction. Thus, comparing the surround-modulation and control conditions should allow us to address the possible confound of long-range luminance responses.

Our protocol, shown in Figure 2.6, included first a behavioral experiment to ensure that induction decreases with surround modulation frequency. As expected, the induction effect indeed decreased with frequency, consistent with findings in the literature. Next, in a block-design fMRI experiment, we collected responses under the surround-modulation, disk-modulation, and control conditions. We compared responses to these conditions in retinotopically identified regions of interest in the primary visual cortex (V1), as well as V2, V3, and V4 (Figure 2.8). If a region is involved in context-dependent lightness processing, we expect to find a strong response under the surround-modulation condition, approximately equal to that under the disk-modulation condition at low frequencies but not at high frequencies. Further, we expect surround-modulation condition responses to be higher than the control condition at low frequencies. If, however, any of these expectations are not satisfied, for example, if the responses under the surround-modulation condition do not decrease with frequency or if the responses at low frequencies do not differ between the surround-modulation and control conditions, that would call into question the involvement of that area in context-dependent lightness processing.

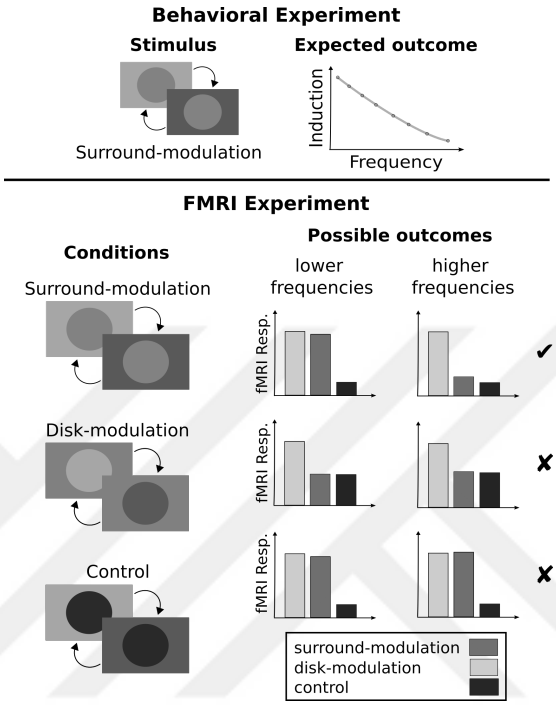

Figure 2.6: First, in a behavioral experiment, we measured the magnitude of induction using the dynamic version of SLI. Based on the results in the literature, we expect that induction decreases with the temporal frequency of the surround modulation. Next, in an fMRI experiment, we measured the fMRI responses under three stimulus conditions and at different modulation frequencies. The 'surround modulation' condition is similar to the stimulus in the behavioral experiment, which leads to the dynamic SLI effect at low but not high frequencies. In the 'disk-modulation' condition, the surround remains constant, and the disk luminance is modulated. In the 'control' condition, the surround is modulated as in the surround-modulation condition, but the disk is black, leading to no dynamic SLI effect. The right panel shows the possible outcomes. If a cortical region is involved in context-dependent lightness processing, the fMRI responses under the surround-modulation condition at low frequencies would be larger than those at high frequencies. Further, the responses under the surround-modulation condition would be larger than those under the control condition at low frequencies. The first row shows this possible outcome. Any outcome inconsistent with this pattern would call into question the involvement of that area in context-dependent lightness processing. The next two rows show two such possible outcomes. The first one would be from a region whose activity is not correlated with lightness and, at best, directly responds to luminance modulations in the surround. The bottom row shows the outcome if the region responds to the border contrast variation, not to the lightness variations in the disk.

2.4 Materials and Methods

2.4.1 Participants

Seven participants participated in the behavioral experiment, and a different set of ten participants participated in the fMRI experiment. All participants had normal or corrected-to-normal vision, and they gave their written informed consent before the experiments. Experimental procedures were approved by the Research Ethics Committee of Bilkent University.

2.4.2 Behavioral Experiment

This experiment was conducted to measure the magnitude of dynamic SLI with different surround modulation frequencies and at different eccentricities within the central disk.

2.4.2.1 Experimental Setup and Stimulus

Visual stimuli were presented on a color reference LCD screen in an otherwise dark room (Eizo CG2730, size: 27 inches, pixel resolution: 2560 x 1440, refresh rate: 60 Hz). A luminance look-up table was prepared through direct measurement to characterize the monitor. The maximum and minimum achromatic luminance of the screen was 100 and 0.015 cd/m^2, respectively. Participants were seated 60 cm from the screen with their head movements constrained using a head-and-chin rest. Participant responses were collected through a standard computer keyboard. Psychtoolbox functions [131] on MATLAB (version 2018a, MathWorks Inc., New York, NY, USA) were used to program the experiments on a Dell 3630 Workstation running Ubuntu Linux (version 18.04).

Figure 2.7 shows the details of the experimental protocol. The basic stimulus

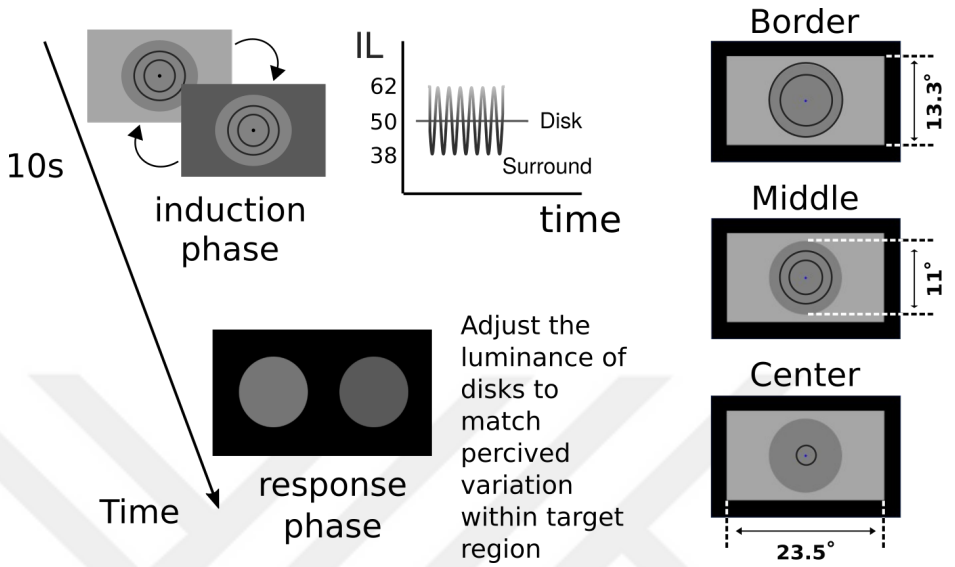

Figure 2.7: A trial in the behavioral experiment was comprised of an induction and a response phase. During the induction phase, we presented the dynamic SLI stimulus at different surround-modulation frequencies ranging from 0.25 to 16 Hz. Participants were instructed to fixate on the fixation mark and judge the perceived lightness variation in a target region marked by thin concentring circles. The right column shows the three target regions, border (8°-11°), middle (4°-7°), and center (0°-3°).

consisted of an achromatic disk with a diameter of 11 degrees of visual angle, embedded in an achromatic rectangular surround subtending 23.54 by 13.37 degrees of visual angle. The rest of the screen was kept at minimum luminance. The luminance of the disk and surround was defined in terms of instrument luminance (IL)

$$IL = \frac{L - Lmin}{Lmax - Lmin} \times 100\%, \tag{2.1}$$

where $Lmin$ and $Lmax$ are the minimum and maximum luminance values of the screen, respectively, and L is the luminance of the stimulus (disk or surround). The luminance of the disk was kept constant at 50% IL, and the luminance of the surround varied between 62% and 38% IL, as described in more detail in the next section.

2.4.2.2 Procedure

An experimental session started with a three-second fixation period during which the luminances of the disk (50% IL) and surround (62% IL) were constant. After this, experimental trials started. Each trial consisted of an induction and a response phase. During the induction phase, the luminance of the disk remained constant (50% IL), and the surround luminance was sinusoidally modulated between 62% (starting value) and 38% IL, with one of 8 frequencies, 0.25, 0.5, 1, 2, 4, 6, 8, and 16 Hz. This produced a perceptual modulation in the disk out of phase with surround modulation, particularly for the lower frequencies. During both the fixation and the induction periods, participants were asked to fixate on a small blue mark at the center of the disk. In each trial, participants judged the perceived luminance variation within one of three target regions of the disk, center region (0°-3°), middle region (4°-7°), and border region (8°-11°).

After the induction phase ended, the response phase started, during which two luminance-adjustable disks were presented side by side on a black background. Participants were asked to adjust the luminance of the disks such that the difference between them corresponds to the difference between the perceived maximum and minimum luminance of the target region within the disk in the stimulation phase. The luminance of the disks could be adjusted by pressing the up and down arrow keys, and the trial was finalized by pressing the space bar, after which the next trial started. The luminance difference between the two disks was taken as the magnitude of SLI.

Overall, there were 24 conditions (8 frequency x 3 target regions), and each condition was repeated ten times, making a total of 240 trials. Trials were randomized across participants. Data were acquired in 3 sessions from each participant.

2.4.2.3 Behavioral data analysis

To analyze the behavioral data, we ran Linear Mixed Models (LMMs) using the 'lme4' package [132] in RStudio [133]. The magnitude of the dynamic simultaneous lightness induction (SLI) in units of percent instrument luminance (% IL) was taken as the dependent variable. Eccentricity, frequency, and their interactions were included as fixed effects, and the between-participant variance was estimated using a random intercept in the model. The model was specified as

$$SLI \sim Eccentricity \times Frequency + (1|\, participant\, ID). \tag{2.2}$$

To calculate the significance of the fixed effects, we used the 'lmerTest' package [134], which uses Satterthwaite's method to estimate the degrees of freedom and generate p-values for linear mixed models. The assumptions of linear mixed models were met: The normality assumption of residuals was assessed by examination of the QQ-plot and Shapiro-Wilk test, and the homogeneity of residual variance assumption was assessed by Levene's test. The package 'emmeans' [135] was used for obtaining the contrasts between estimated marginal means adjusted for multiple comparisons using Tukey's test. The package 'ggplot2' was used for data visualization [136].

2.4.3 fMRI Experiment

2.4.3.1 Experimental Setup and Data Acquisition

A 3 Tesla Siemens Trio MR scanner (AG, Erlangen, Germany) with a 32-channel phase-array head coil was used to acquire the MR data at the National Magnetic Resonance Research Center (UMRAM), Bilkent University. Each participant underwent multiple MR runs in a session, including one run for high-resolution anatomical imaging (T1-weighted 3D MPRAGE sequence, spatial resolution: 1 mm^3 isotropic, TR: 2600 ms, TE: 2.92 ms, flip angle: 12°, number of slices: 176), and a total of 8 functional runs, one for retinotopic mapping, one for functional

region of interest (ROI) identification, and six for experimental conditions (T2*-weighted EPI sequence, spatial resolution: 3x3x3 mm^3, TR: 2000 ms, TE: 35 ms, flip angle: 75°, number of slices: 30, slice orientation: parallel to calcarine sulcus). An MR-safe 32-inch LCD monitor (T-32, Troyka Med A.S., Ankara, Türkiye) was used to present the visual stimuli. The maximum and minimum achromatic luminance of the screen was 79.6 and 0.0024 cd/m^2, respectively. The participants lay supine in the scanner and viewed the screen using a mirror attached to the head coil above their eyes, with a total viewing distance of about 172 cm.

2.4.3.2 MR Data Preprocessing and Analysis Software

MR data were pre-processed and analyzed using FSL and Freesurfer [137, 138]. Preprocessing steps included BET brain extraction applied on T1-weighted anatomical images to remove non-brain tissue. For functional T2*-weighted images, preprocessing steps included MCFLIRT motion correction, BET brain extraction, and high-pass temporal filtering. Functional images were aligned to the individual's T1-weighted anatomical image and registered to standard T1 MNI152 brain resampled at 1 mm cubic voxels. For each individual participant, the 3D cortical surface was constructed from anatomical images using Freesurfer for visualizing statistical maps, delineating early visual areas, and identifying regions of interest.

2.4.3.3 Retinotopic Mapping and ROI identification

Horizontal and vertical wedges mapped with flickering checkerboard patterns were used in a block-design fMRI run to delineate the early visual areas V1, V2, V3, and V4 of each participant following the standard procedures in literature (Engel et al., 1997; Sereno et al., 1995; Slotnick et al., 2003). Stimulation code was written in the Java programming platform by us. The data were preprocessed as described above, and a general linear model (GLM) was applied using FSL. The resulting parametric maps were projected onto the 3D reconstruction of

the participant's cortex in native space. Then, Freesurfer's drawing tools on its Freeview program were used to draw the boundaries between the early visual areas.

ROIs were identified within each visual area using a block-design localizer run. Stimulation code was written by us using the Psychtoolbox libraries [131] on MATLAB (version 2018a, MathWorks Inc., New York, NY, USA). The basic stimuli were pairs of arcs mapped with flickering high-contrast checkerboard patterns (Figure 2.8). During the localizer run, in each 16-sec block, one of four pairs of arcs was presented to the participants at different eccentricities. The positions of the arcs corresponded to the center (1.5 to 2.5 degrees eccentricity), middle (5 to 6.5 degrees eccentricity), border (9 to 10.5 degrees eccentricity), and surround (14 to 16 degrees eccentricity) of the experimental disk. The data were preprocessed, and a general linear model (GLM) was applied using FSL. The resulting parametric maps were projected onto the 3D reconstruction of the participant's cortex in native space. Based on this and the retinotopic mapping, ROIs within each early visual area were identified, e.g., V1 middle ROI, etc.

2.4.3.4 Experimental Design, Stimuli, and Data Analyses

fMRI responses to experimental stimuli were collected using a block-design paradigm. Three conditions, surround-modulation, disk-modulation, and control, were tested in different runs (Figure 2.6). There were two runs per condition. In all three conditions, the basic stimulus was composed of a gray-scale rectangle (surround) and a central disk with the same dimensions as in the behavioral experiment. In the surround-modulation condition, as in the behavioral experiment, the luminance of the disk was kept constant (27.6% IL), and the surround luminance was modulated sinusoidally (between 20.2% and 35.6% IL). In the disk-modulation condition, the luminance of the disk was sinusoidally modulated while the luminance of the surround was kept constant (27.6% IL). Here, the modulation of the disk followed the same temporal sinusoidal pattern as the surround in the surround-modulation condition (between 20.2% and 35.6% IL). The stimulus in the control condition was the same as in the surround-modulation

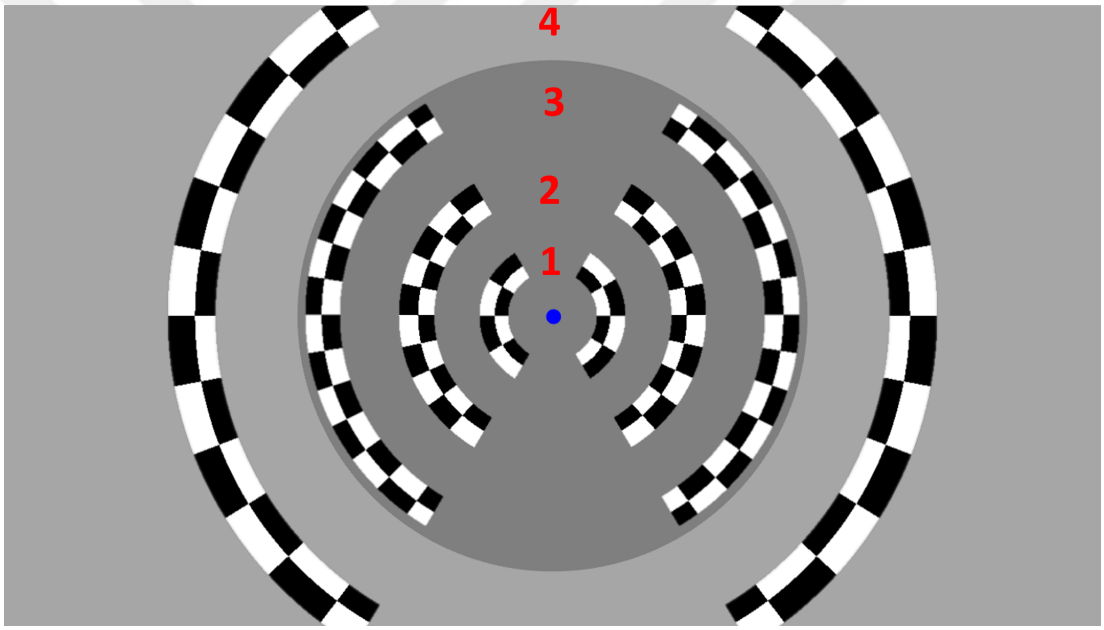

Figure 2.8: Functional localizer stimuli. High-contrast counter-phase flickering checkerboard patterns were presented in pairs of arcs on a uniform gray background (arcs are overlaid on the experimental stimulus in this figure to provide a visual reference to the readers). Arcs 1, 2, 3, and 4 were used to localize the center (1.5°- 2.5°), middle (5°- 6.5°), border (9°- 10.5°) and surround (14°- 16°) ROIs in V1, V2, and V3. For V4, only border and center ROIs were extracted.

condition, except that the disk was black (0% IL). Four sinusoidal modulation frequencies, 1, 2, 4, and 8 Hz, were tested in all conditions.

Each run started with a static block where a static screen composed of the disk, the surround, and the fixation mark was presented for a period of 12 seconds to allow for the fMRI signal to stabilize. This was followed by a 12-second test block in which the luminance of the surround (surround-modulation and control conditions) or the disk (disk-modulation condition) was modulated at one of the four modulation frequencies. The dynamic block was followed by another static block, and this pattern was repeated during the run such that each frequency condition was presented four times. Frequency conditions were pseudo-randomized and counter-balanced to rule out any order effects. Throughout the run, the fixation mark remained visible in all stimulation blocks. The default color of the fixation mark was blue, and it changed to red or yellow every 1500-2500 ms for 100 ms. Participants were asked to keep their fixation at the mark and report any changes in its color using an MR-safe response box (Fiber Optic Response Devices Package 904, Current Designs). The run was concluded with a 12-second static block. The total duration of an experimental run was 396 seconds.

Using FSL, we first ran a whole-brain GLM analysis. Using the functional ROIs as masks and the Featquerry tool of FSL, we extracted the GLM beta-weights from each ROI and treated them as fMRI response amplitudes. Further statistical analyses were performed on the extracted fMRI response amplitudes in RStudio [133].

For V1, we ran Linear Mixed Models (LMMs) using the 'lme4' package [132] where the magnitude of the fMRI response amplitude was taken as the dependent variable. Condition (three levels; disk modulation, surround modulation, and control) Eccentricity (four levels; center, middle, border and surround), frequency (four levels; 1, 2, 4, and 8 Hz), and all the subsequent interactions were included as fixed effects. The between-participant variance was estimated using a random intercept in the model. The model was specified as

$$fMRI\,response \sim Condition \times Eccentricity \times Frequency + (1|participant\,ID). \tag{2.3}$$

For each of V2, V3, and V4, a separate model was specified. We report only the results in the center ROIs because responses in those ROIs are least affected by the border contrast variations, as explained further in the Results section. Thus, eccentricity is not included as a fixed effect factor in the models. The rest of the specifications were the same as the model for V1:

$$fMRI\ response \sim\ Condition \times Frequency + (1|\ participant\ ID) \qquad (2.4)$$

As in the behavioral analysis, we used the 'lmerTest' package [134] to calculate the significance of the fixed effects. We found that the assumptions of linear mixed models were met. The normality assumption of residuals was assessed by examination of QQ-plots and Shapiro-Wilk test. The homogeneity of residual variance assumption was assessed by Levene's test. The package 'emmeans' [135] was used for obtaining the contrasts between estimated marginal means adjusted for multiple comparisons using Tukey's test.

2.5 Results

2.5.1 Behavioral experiment

The left panel of Figure 2.9 shows the means of the magnitude of the dynamic simultaneous lightness induction (SLI) as a function of surround modulation frequency. Consistent with previous literature, we found that SLI decreases with surround modulation frequency. This trend is observed in all eccentricities tested. Linear mixed model (LMM) analysis on the SLI magnitudes showed a significant main effect of frequency ($F(7) = 116.3$; $p<0.001$), a significant main effect of eccentricity ($F(2) = 183.4713$; $p<0.001$), as well as a significant interaction between eccentricity and frequency ($F(14) = 3.0559$; $p<0.001$). As can be observed in Figure 2.9, the induction is significantly stronger near the border between the disk and the surround. This is also reflected in post hoc pairwise comparisons of estimated marginal means among eccentricity conditions, which show that the magnitude of SLI is significantly different between all eccentricity conditions

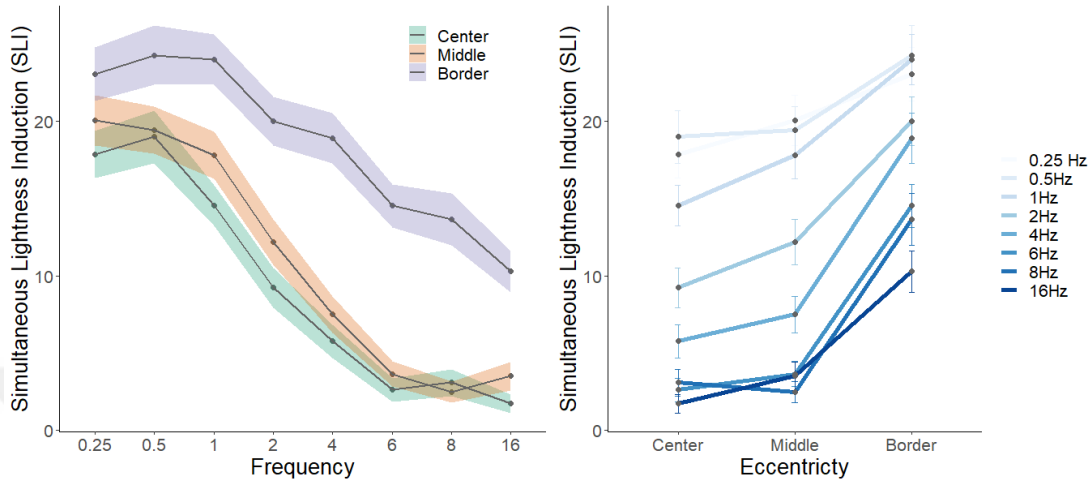

Figure 2.9: Behavioral results. Left panel: Group average of the dynamic simultaneous lightness induction (SLI) as a function of surround modulation frequency for each eccentricity condition. SLI is higher at the outer eccentricity, which is closest to the border between the disk and the surround. In the middle and center eccentricities, SLI is highest at lower surround modulation frequencies, decreases with frequency, and nearly disappears for the highest frequencies. Shaded regions show the standard error around the mean. Right Panel: Group average of the dynamic simultaneous lightness induction (SLI) as a function of eccentricity for each frequency condition.

($p<0.01$), with the border region producing the highest estimated marginal mean (18.56) in comparison to the middle (10.81) and center (9.22). These results show that for our stimulus configuration, lightness induction at the center of the disk is stronger for lower modulation frequencies, decreases for higher frequencies, and almost disappears at 8 and 16 Hz. Based on these results, we chose four surround modulation frequencies, 1, 2, 4, and 8 Hz, to be tested in the fMRI experiment.

2.5.2 fMRI experiment

2.5.2.1 V1 Results

Figure 2.10 shows the fMRI responses from the primary visual cortex (V1) averaged over ten participants under the disk-modulation and surround-modulation

conditions. Linear Mixed Models (LMMs) analyses with conditions (disk-modulation, surround-modulation, and control), frequency (1, 2, 4, and 8 Hz), and eccentricity (center, middle, border, and surround) as fixed effect factors revealed a significant main effect of eccentricity ($F(3) = 28.3$; $p < 0.001$), a significant main effect of frequency ($F(3) = 27.7$; $p < 0.001$), a significant main effect of condition ($F(2) = 55.8$; $p < 0.001$), and a significant interaction between eccentricity and condition ($F(6) = 8.2$; $p < 0.001$). The responses in the border ROI under the surround-modulation and disk-modulation conditions are not different ($p = 0.8352$), and they both increase with frequency (Figure 2.10, Border ROI panel). This shows that the responses in the border ROI are dominated by the contrast variation at the border between the disk and the surround. We further note that the fMRI responses are smaller in the center and middle ROIs compared to the border ROI at all frequencies. This reduction appears systematic and depends on the distance between the border and the ROI. In light of these observations, we focus only on the center ROI in our analyses, as it is likely to be least affected by the dynamic border contrast and best represents the neural activity related to lightness processing.

Figure 2.11 left panel shows averaged fMRI responses in the center ROI. Posthoc pairwise comparison of estimated marginal means from LMM revealed a significant difference between the surround-modulation and disk-modulation conditions only at 8 Hz ($p < 0.05$). To further aid visualization, we computed the differences between the responses under the two conditions (Figure 2.11, right panel). Those results show that relative responses to surround modulation, just as the behavioral SLI effect, decrease with frequency. This provides evidence that the responses reflect neural activity related to the context-dependent SLI effect.

Figure 2.12 shows the fMRI responses under the control condition in the center ROI, together with the responses under the surround-modulation and disk-modulation conditions. At 1 and 2 Hz, fMRI responses under the surround-modulation condition are stronger than those under the control condition and similar to those under the disk-modulation condition. But at higher frequencies, particularly at 8 Hz, fMRI responses under the surround-modulation condition

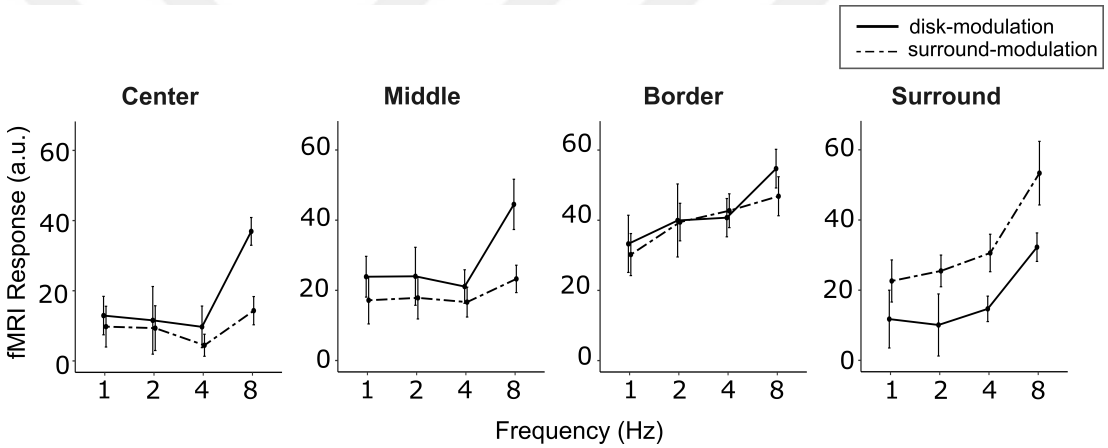

Figure 2.10: V1 group average of fMRI responses as a function of frequency in all four ROIs. fMRI responses near the border between the disk and surround are the same for the surround-modulation and disk-modulation conditions for all frequencies (Border ROI plot). This shows that those responses are likely caused by the contrast modulation at the border. Whereas in the central and middle ROIs, fMRI responses under the surround-modulation and disk-modulation conditions remain similar only at 1, 2, and 4Hz, but diverge at 8Hz (Central ROI and Middle ROI plots), supporting that V1 responses in those ROIs are correlated with SLI, not simply the contrast modulation at the border. Error bars: SEM.

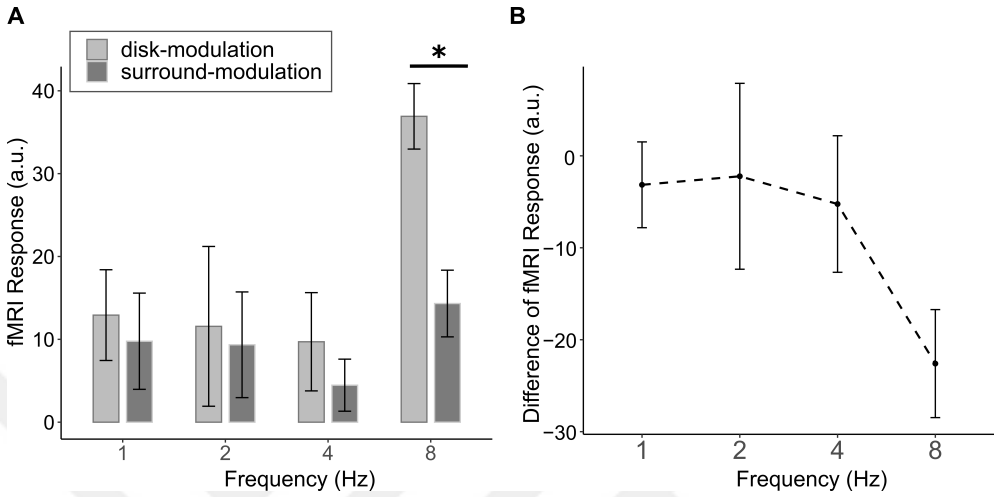

Figure 2.11: **A**. fMRI responses in the center ROI under surround-modulation and disk-modulation conditions, replotted from Figure 2.10 center ROI plot. * indicate a significant difference under the pairwise comparison of estimated marginal means. **B**. The difference between responses under the two conditions, surround-modulation minus disk-modulation, as a function of frequency.

are close to those under the control condition and less than those under the disk-modulation condition. These results further support that fMRI responses under the surround-modulation condition reflect neural activity related to perceptual lightness induction, not long-range luminance responses (as described in Figure 2.6).

2.5.2.2 Extrastriate Areas

Figure 2.13 shows averaged fMRI responses in the center ROIs of visual areas V2, V3, and V4. Linear Mixed Models (LMMs) analysis, which was performed on each visual area separately, with conditions (disk-modulation, surround-modulation, and control) and frequency (1, 2, 4, and 8 Hz) as fixed effect factors, revealed a significant main effect of condition in all areas ($p < 0.001$), a main effect of frequency only in V4 ($p < 0.05$), and no significant interactions between condition and frequency. Posthoc pairwise comparison of estimated marginal means revealed that fMRI responses under the control condition are significantly different from those

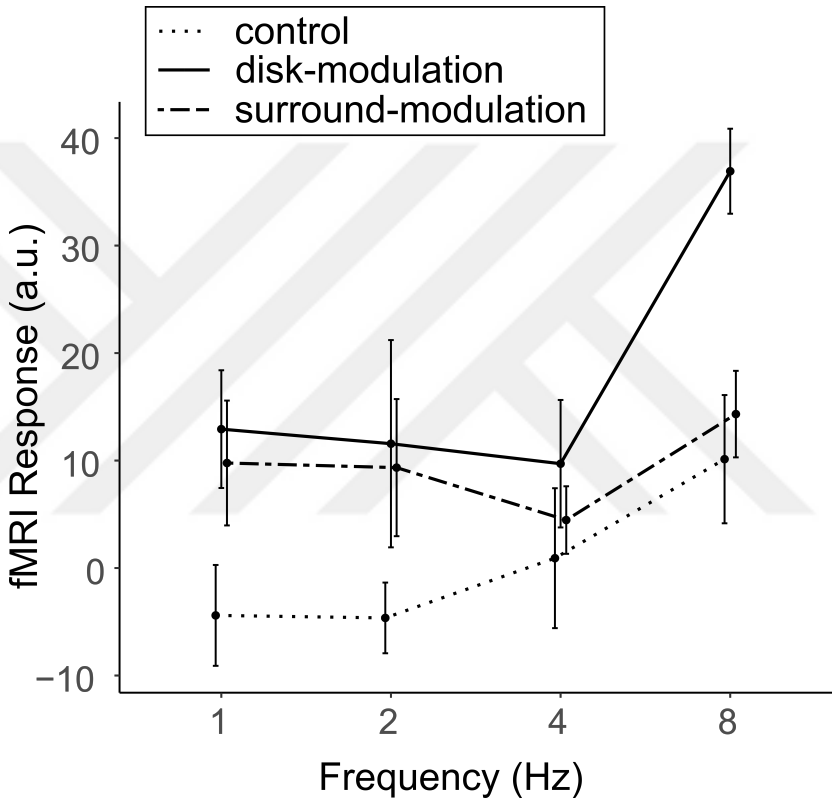

Figure 2.12: fMRI responses in the V1 center ROI under the surround-modulation, disk-modulation, and control conditions. At low frequencies, responses under the surround-modulation condition are approximately the same as those under the disk-modulation condition and larger than those under the control condition. Whereas, at high frequencies, especially at 8Hz, this relation reverses. At 8 Hz, the responses under the surround-modulation condition are approximately the same as those under the control condition and lower than those under the disk-modulation condition. This is exactly what would be expected from an area whose activity is correlated with the behavioral lightness induction effect, as depicted in Figure 2.6. Error bars represent SEM.

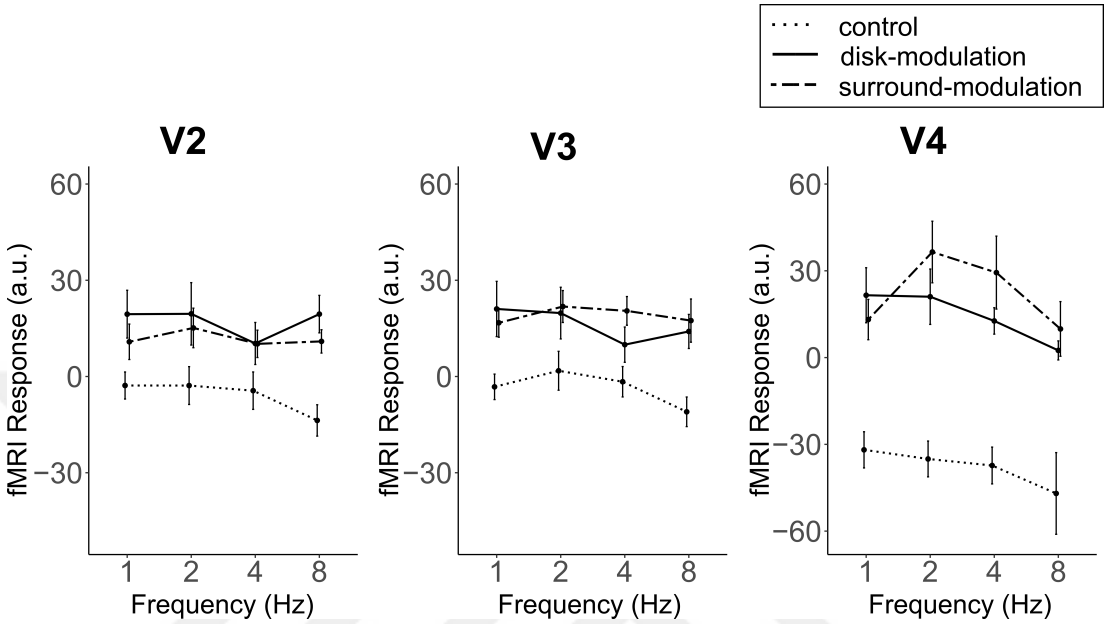

Figure 2.13: fMRI responses in V2, V3, and V4 center ROIs. Because it was not possible to localize the center ROIs in all participants, the data were averaged across 8 participants in V2 and V3 and across 7 participants in V4. Responses under the disk-modulation and surround-modulation conditions did not differ in any area, and they were larger than those under the control condition. The pattern of results suggests that the fMRI response is dominated by contrast variations at the border, as depicted in Figure 2.6. Error bars represent SEM.

under the disk-modulation and surround-modulation conditions. However, there was no significant difference between surround-modulation and disk-modulation conditions in any area. This pattern of results resembles the last possible outcome depicted in Figure 2.6, and suggests that the responses are dominated by contrast variation at the border, thus, it does not provide strong evidence that V2, V3, and V4 are involved in context-dependent lightness processing. We discuss these findings in the next section.

2.6 Discussion

2.6.1 Spatiotemporal properties of simultaneous lightness induction

Our behavioral study demonstrated a spatial non-uniformity in the magnitude of lightness induction across the disk for all frequency conditions. The magnitude of perceived lightness is the strongest near the border of the surround and disk, and decreases towards the center. This finding is in line with the results of previous studies [79, 81, 82]. Davidson [82] and Cornsweet [81] demonstrated such spatial non-uniformities in static setups. Pereverzeva et al. [79] used a dynamic stimulus similar to what we used and reported that the magnitude of perceived lightness decreases as the distance from the border increases. However, they studied this spatial nonuniformity at the surround modulation frequency of 1 Hz. The current study investigates spatial nonuniformity at multiple frequencies. These findings are discussed in light of the temporal properties of lightness induction below.

We found, consistent with the findings in the literature, that the dynamic simultaneous lightness induction effect is stronger at lower surround modulation frequencies, decreases with increasing frequency, and nearly vanishes at very high frequencies [23, 59, 79, 86]. The main goal of our behavioral results was to choose the frequencies for the fMRI experiment where the magnitude of induction appears to vary the most. So, we chose four frequencies from 1 to 8 Hz based on the results. However, we also observed an interesting trend of temporal cut-off when eccentricity is taken into account. The frequency cut-off for lightness induction that is being reported in the literature is approximately 2-4 Hz [23, 59, 61, 86]. We observed that the cut-off in the middle and center of the disk is consistent with the reported cut-off from previous studies, but the magnitude of induction is quite higher than the reported cut-off at the border. What causes a higher frequency cut-off at the border, and why is there a very big increase in induction as we move from the middle to the border of the disk for higher frequencies? These questions might raise a possibility of a more complex percept

happening at the border than at the central position of the disk. De Valois et al. [86] studied the magnitude of induction as a function of the frequency of surround modulation and have shown similar results to what we found. They studied the frequencies up to 8 Hz. They report that the central region appears to change the lightness when the surround is modulated at low frequencies, but at high frequencies, the region appears to flicker. Hence, the high magnitudes of induction we get for 8 and 16 Hz might reflect this flicker. Moreover, some studies report the percept of flicker halo at higher frequencies at the border, which might also be a confounding factor here [139]. However, our behavioral task did not aim to disentangle other percepts from the lightness perception, so these effects should be investigated further.

2.6.2 Neural correlates of spatiotemporal properties of simultaneous lightness induction

Our results show that fMRI responses in the primary visual cortex (V1) correlate with perceived lightness in the dynamic simultaneous lightness induction (SLI) effect. This outcome supports that V1 is involved in context-dependent lightness processing. We did not, however, find strong fMRI evidence that extrastriate visual areas, V2, V3, and V4, are involved in context-dependent lightness processing.

2.6.2.1 Role of V1 in context-dependent lightness perception

In the fMRI experiment, we measured responses from visual areas while observers viewed the dynamic SLI stimulus. We found that V1 responses in the region corresponding to the center of the disk reflect the pattern in the behavioral results: lightness-related responses are stronger at low frequencies than those at high frequencies. This finding is in line with previous animal electrophysiological studies [24, 88, 110, 140, 141] as well as human fMRI studies [113, 26, 29, 112, 128], and support that the activity of V1 correlates with context-dependent lightness

perception.

Is the response in V1 merely edge response? To address the possible confounding effect of border contrast variations (e.g., see Cornelissen et al. [30]), we systematically assessed how the fMRI responses varied as the modulation frequency increased. We performed this analysis in functionally defined regions of interest (ROIs) corresponding to different parts of the disk. If the responses were primarily driven by the border contrast, not induced lightness, then we would expect them to increase with frequency [142, 143]. But that was not the case for all ROIs. In the ROI closest to the border of the disk (border ROI), the responses indeed increased with frequency, indicating that they are largely driven by the border contrast. fMRI responses to the parts of the disk further from the border (central and middle ROIs), however, showed little or no increase with frequency. This pattern speaks against a border contrast explanation.

To further assess how activity varies with frequency, we computed the responses under the surround-modulation condition relative to those under the disk-modulation condition, where the surround remains static, and the disk is modulated in luminance. Because the border contrast variation was the same across the two conditions, this analysis allowed us to compare the responses to luminance and induced lightness variations directly. At low frequencies, where there is a strong behavioral effect, the responses under the surround-modulation and disk-modulation conditions were nearly equal in the ROI corresponding to the center of the disk. But at the highest frequency, where induction disappears, the responses to the surround-modulation condition were significantly less than those to the disk-modulation condition. In other words, relative responses under the surround-modulation condition decreased with frequency, in line with the behavioral results (Figure 2.11).

We noted that the fMRI responses under the surround-modulation condition might not be related to context-dependent lightness but merely reflect the activity of neurons that respond to the surround modulation or the overall illumination in the environment. A comparison of responses under the surround-modulation and control conditions rules out this possibility. Under both conditions, the

surround is modulated with the same temporal characteristics. But only under the surround-modulation condition does the disk's lightness vary. If the fMRI responses merely reflect the neural activity in response to the surround modulation, they should be the same under both conditions at all frequencies. But this is not the case. At the highest frequency, where there is no lightness variation under the surround-modulation condition, fMRI responses are almost equal under the surround-modulation and control conditions (Figure 2.12). This suggests that when perceptually there is no induction, the fMRI response might indeed be driven by the surround variation alone. But at low frequencies, the fMRI responses under the surround-modulation condition are stronger than those under the control condition. In other words, when there is lightness induction, the fMRI responses are stronger compared to when there is no lightness induction. This pattern argues against a long-range response explanation.

2.6.2.2 Role of higher visual areas in context-dependent lightness perception

Previous studies have found evidence that extrastriate areas V2, V3, and V4 are involved in context-dependent lightness processing [113, 77, 115, 116, 118, 119, 128, 144, 145, 146]. Here we found that, in V2, V3, and V4 center ROIs, fMRI responses under the disk-modulation and surround modulation conditions were nearly the same, and they were both larger than those under the control condition at both low and high frequencies. At low frequencies, this is the expected pattern of results in an area involved in lightness processing. The pattern at high frequencies, however, suggests that the responses may be caused by the border contrast variation and raises doubts that they are related to the lightness induction (the last possible outcome depicted in Figure 2.6). Thus, our results provide limited support that extrastriate areas V2, V3, and V4 are involved in context-dependent lightness processing.

This outcome could be the result of fMRI's limitation as a technique to resolve neural activity. Specifically, a measurement from an fMRI unit, voxel, reflects an aggregate and indirect response from hundreds of thousands of neurons that

may have heterogeneous response characteristics. One of those characteristics is the (population) receptive field size, which is typically larger in extrastriate areas compared to V1 [147, 148]. Even though we have carefully identified voxels that, on average, respond more strongly to our flickering checkerboard localizer near the center of the disk, these voxels may still harbor neurons that have relatively large RF sizes and respond to the contrast variation at the border. Subtle variations with frequency in the lightness-related neural activity could be buried under those long-range activations. Thus, our fMRI results do not unequivocally prove that extrastriate areas are not involved in context-dependent lightness processing.

Interestingly, unlike in V1, fMRI responses in extrastriate areas do not increase with increasing frequency under the disk-modulation condition. In V4, we even observe a decrease in responses with frequency. Previous fMRI studies investigating responses of the visual cortex to chromatic or/and achromatic temporal modulations reported that V4 shows a low pass temporal dependence on frequency [149, 150]. This reduced temporal resolution in V4 can be attributed to its involvement in form perception and object recognition, where temporal information may play a less critical role. Further, V4 plays a role in surface perception, and its responses may show dissociation with those of V1 if the stimulus does not appear to be a surface [126]. Thus, an alternative explanation for our results could be that at high frequencies, V4 may not treat the stimulus as a surface anymore, leading to a decrease in its responses with frequency.

2.6.3 The neural mechanism of context-dependent lightness perception

The current study provides evidence for the involvement of cortical mechanisms for processing the of context dependent lightness induction in line with previous animal single-cell physiological [24, 88, 110, 112, 115, 140], human electrophysiology [151] and fMRI studies [29, 113].

One of the potential cortical mechanisms is the filling-in mechanism as described in the introduction section. According to the filling-in mechanism, the propagation of a signal from the edges should take some time. At the frequencies of luminance modulation of surround at which the propagation speed of filling-in signal is less than the modulation frequencies, there should be a drop off in filling-in signal. Hence, the behavioral part of the current study provides evidence for filling-in mechanism by showing that there is a decrease in the magnitude of perceived lightness induction at higher frequencies in line with previous studies [86, 23, 79]. In addition, we show that the neural correlates of this decreasing trend in the magnitude of lightness induction are also reflected in cortical activity, hence providing a possibility of filling-in the mechanism to occur in V1. Another potential cortical mechanisms is the multiscale spatial filtering approach which has been discussed in the introduction section. Such a model predicts lightness induction to be an instantaneous process [48]. Hence, the results of the current study cannot be explained by the multiscale spatial filtering approach.

One of the prominent criticism of filling-in account comes from Blakeslee and McCourt [62], who argue against lightness induction being a slow process. Using grating induction, they showed that there was no delay in the perception of induced grating and actual grating. They used this finding to provide evidence that it is an instantaneous process. However, while comparing the perception of actual and induced grating, Blakeslee and McCourt [62] did not take into account the fact that if the same mechanism lies behind the induced and actual grating perception, no delay should be expected between both. Another criticism of filling-in account is that the data from the spatial properties of lightness induction show a spatial nonuniformity in the magnitude of induction. A simple filling-in model is not able to predict such inhomogeneity. Instead, this property is consistent with multiscale spatial filtering accounts. The results of the current study do show that induction is spatially non-uniform. However, if we look at the lightness induction magnitudes at the border region, it shows a more complex percept happening at the border, which might be a confounding factor leading to this non-uniformity. This needs to be further investigated in future studies.

In conclusion, the current study provides evidence for filling-in mechanism,

but the spatial non-uniformity found in lightness induction magnitudes should be investigated further while controlling for possible confounds. We also do not rule out the possibility of the role of top-down influences from higher-level areas in addition to low-level filling-in mechanism.

2.6.4 Conclusion

In summary, our findings provide further evidence that the activity of V1 correlates with context-dependent lightness, not only with luminance. This suggests that V1 is involved in context-dependent lightness processing. From a broader perspective, this result adds to the body of results in the literature, which suggests that the activity of V1 does not merely correlate with the physical properties of the retinal image but also reflects perceived qualities. However, what is the exact mechanism behind this is still not known.

2.7 Future Directions

2.7.1 An effective connectivity study

This thesis aimed to investigate the neural correlates of contextual processing of lightness. The spatiotemporal properties of context-dependent processing do provide insights into the mechanisms governing this percept, but we did not directly aim to look into the mechanisms governing this percept. As mentioned in the introduction section, filling-in, and multiscale spatial filtering models are two prominent mechanisms supporting low-level accounts. Filling-in mechanism well explains the lightness stimuli that depend on high spatial frequency, such as the Cornsweet illusion or the Koffka rings. A multiscale approach cannot explain these stimuli since small edges do not make much difference in terms of energy in an image. But stimuli like grating induction can be greatly explained by a multiscale approach where the lightness of a uniform region appears non-uniform [106]. Using SLI as a stimulus, the literature has produced mixed evidence for both of these mechanisms. Investigating effective connectivity or causality amount regions corresponding to different eccentricities within an induced region can be useful to find whether a filling-in mechanism accounts for lightness processing or not. As previously explained, the filling-in mechanism propagates signal from the edges to the center of the disk/induced region. If this is really the case, then we should see a delay between the BOLD signal corresponding to the border and the center of the region to be filled in. In other words, the signal from the central region must be caused by the signal at the border. To achieve this, one can extract the BOLD signal corresponding to the edges and center of the disk and use Granger causality or DCM (Dynamic causal Modelling) to model the causality. If filling-in is the potential mechanism behind lightness processing, there should be a causal relation from border ROI to central ROI, as shown in Figure 2.14.

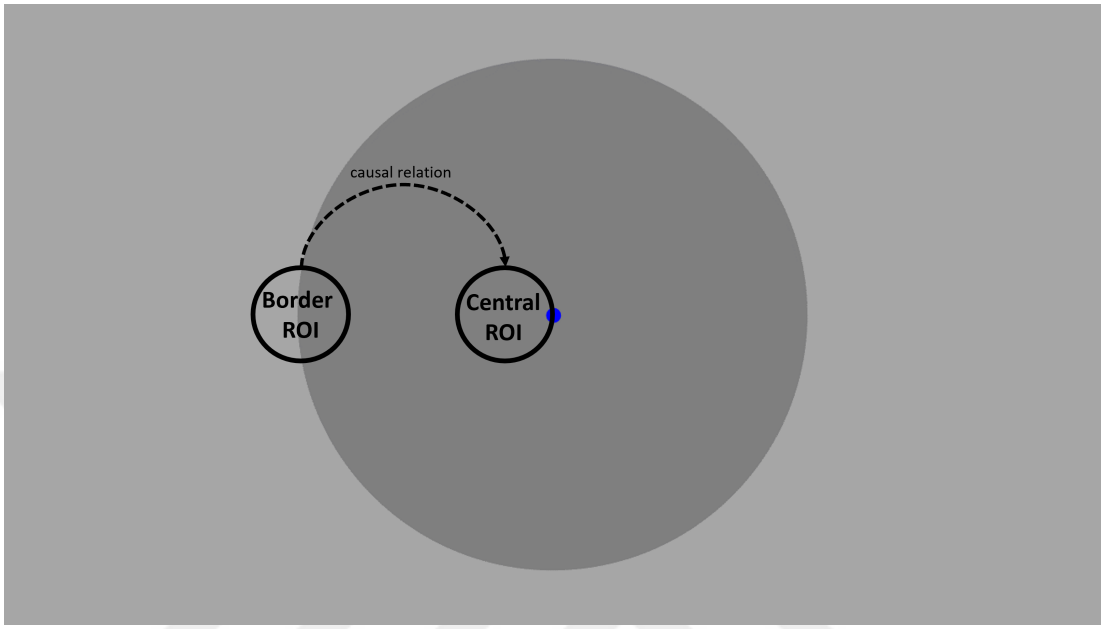

Figure 2.14: Causality study: There should be a causal relation from border ROI to central ROI if lightness is calculated indirectly from edges at the center of the disk (filling-in mechanism). However, if this is not the case, there should not be a causal relation between the border and central ROI.

2.7.2 Role of higher areas

Even though the results of the current study provide evidence for the involvement of the primary visual cortex, we do not exclude the possibility that the activity in V1 is modulated by a higher area. Higher areas like lateral occipital complex (LOC) and intraparietal sulcus (IP) have been shown to be involved in lightness processing [77]. The current study did not aim to localize and investigate these areas. Examining the activity of these areas corresponding to the stimulus used in this study can also be helpful.

2.7.3 Genralizibility across different stimuli

One reason for the inconsistent findings regarding mechanisms of context-dependent lightness processing among different studies is the use of different stimuli. A wide range of stimuli demonstrate the effect of context on lightness,

which vary so much in their configuration. It is very difficult to compare the perceptual magnitude of contextual lightness perception in these stimuli while controlling for all the confounds. The current study made use of simultaneous lightness induction (SLI) stimulus. Examining the activity of brain regions corresponding to other lightness stimuli in the same way while controlling for contrast, mean luminance, and other factors and then comparing the results across studies can be helpful to at least agree on the brain region involved in contextual lightness processing. Moreover, none of the currently proposed models can explain all of the lightness stimuli demonstrating the effect of context, which calls for the new model to take into account all the limitations of existing models.

Nonetheless, understanding the role of low-level visual areas is essential. Understanding the complex mechanisms and role of higher areas is difficult when the role of low-level areas, which lie at the initial stages of processing, is missed or ignored. The study presented in the current thesis will help to progress the literature on the processing of lightness in context by highlighting the role of the primary visual cortex.

Chapter 3

Effect of expectations on the perception of dynamic material properties

3.1 Introduction

Traditionally, perception was treated as a unidirectional bottom-up process in which information simply flows from the sensory organs to the brain in orderly steps. For instance, in visual perception, light is reflected from objects and enters the retina. This light is detected by the photoreceptors in the retina, which transduce the light energy into electrical signals in neurons. The information from the retina then goes through the optic nerve to the thalamus, which then transfers it to the primary visual cortex and ultimately to extrastriate areas. Simple low-level information like orientation and contours is processed in early cortical areas then as the information goes to higher levels, more complex features like object recognition and identification are processed. This one-way flow of information that starts from the raw data from sensory organs to the perception of three-dimensional views of the world at a higher level is called feedforward or bottom-up processing. This view was influenced by the structural organization

of the primate visual system. Earlier work on feedforward processing was based on the computational model provided by David Mar [152] in his book 'Vision' published in 1982. According to this model, visual processing comprised levels of computation starting from simple local analysis, ultimately giving rise to the global 3D interpretation of the whole visual scene [152]. This model inspired the functional model of the visual system. According to this model, neurons in early visual areas are involved in local 2D interpretations of visual scenes. Inputs from early visual areas combine systematically in higher visual areas to interpret more complex properties of the visual scene [153].

Top-down influences in sensory processing: Initially, one-way feedforward processing of information played a huge role in understanding visual systems' structure, organization, and processing. However, later research showed that a one-way system was insufficient to explain the complexity of vision. Hence, soon it was recognized that the feedforward model of visual processing failed to explain the perception of real-world scenes, especially when the incoming sensory information is noisy and ambiguous [154]. For instance, the one-way feedforward model is not able to segment a complex scene into local shapes with a lot of variations of lighting, occlusions, and shadows [155]. The Feedforward model also fails to explain information processing in situations when information from prior experiences influences the perception of the ambiguous incoming sensory information [156]. Moreover, it has been shown that the information in a visual scene is integrated over large distances which are greater than the size of classical receptive fields of V1 and cannot be explained by the local horizontal connections on which the feedforward model is based [157]. In the meantime, investigations into the structural organization of the visual cortex demonstrated a high proportion of feedback projections in corticocortical connections indicating that the role of top-down processing had been greatly underestimated [158].

Since then, a huge amount of research has been done to investigate how top-down processing affects perception. Two cognitive factors that modulate the top-down and bottom-up integration are attention and expectations [156, 159, 160, 161, 162]. Both attention and expectation are believed to make the processing of complex, ambiguous and noisy incoming sensory information

computationally efficient and faster. The role of attention in perception has been studied extensively [163, 164, 165, 166, 167]. However, the effect of expectations on visual processing has become the subject of focus recently [164, 168]. As this thesis focuses on the effect of expectations, I will discuss top-down influences, particularly in the context of expectations.

What are expectations? Cognitive context includes aspects of cognitive processes like engagement, memory, learning, and knowledge. All of which are interconnected. A major implication of the effects of cognitive context on perception is expectation. But what are expectations?

Imagine you go to someone's house, and they ask you to bring them a plate. You would most likely go to the kitchen and look for a plate in the cabinets. Why would you not go to a bedroom instead to look for a plate? When you think of a plate, you instantly imagine it being in the kitchen or on the dining table. Similarly, if you think of a tree, you imagine a garden or a natural landscape scene in the background. You associate a computer with an office, a lectern stand with a classroom, a painting with a wall in the background, and so on. Objects do not exist in isolation in the real world but are always associated with a scene/context. The repetitive encounter with objects in a particular context leads us to associate objects with the backgrounds or with other objects. These associations are not just limited to the visual domain. For instance, if you see a rose, just by looking at it, you would know the kind of smell it carries. Similarly, you associate females with having higher pitch voices than males. These learned associations set our expectations and are then used to make predictions about forthcoming sensory information. Expectations can be defined systematically as the state of the cognitive system just before the processing of incoming sensory stimuli begins. Expectations provide a cognitive context in the form of prior knowledge of the world and learned associations to predict forthcoming sensory information. Expectations influence perception through top-down processes and help to detect ambiguous, noisy, and incomplete stimuli.

Early work on expectations. The first line of work providing evidence for the existence of a top-down processing scheme in addition to bottom-up in the

form of expectations comes from the work of Biederman [169] on the perception of real-world scenes. He argued that the stimuli used in the perceptual recognition experiments are too simple and do not provide a meaningful context in comparison to even real-world simple scenes. He conducted experiments to show that meaningful context, in addition to peripheral scanning and memory, does affect perception. Similarly, Palmer [170] showed that observers performed better in the identification of objects in an appropriate context than inappropriate context. Davenport and Potter [171] showed that not only is object perception affected by inconsistent background, but the background scene perception is also affected by inconsistent objects. They concluded that object and scene perception occurs interactively. These initial findings were followed by several studies in computer vision and visual cognition that have shown that context facilitates object recognition [172, 173, 174, 175, 176, 177, 178]. Later, research on the influence of expectations expanded to auditory perception, pain perception, memory, and decision-making [179, 180, 181, 182]. Early research on expectations laid a foundation for understanding the influence of top-down processing in perception and cognition. To date, it remains an important area for understanding human cognition and behavior.

3.1.1 Importance of expectations in visual perception

Expectations not only help to detect and recognize ambiguous and noisy stimuli but also help organisms to respond to the environment accurately and efficiently. Hence, expectations are considered to play a significant role in the survival of organisms by modulating a range of cognitive processes, from sensory perception to motor responses. However, under the scope of this thesis, I discuss the importance of expectations in the visual domain below.

1) Expectations help us recognize the objects in the surrounding quickly and robustly. Some of the earliest studies have demonstrated the importance of expectations in object recognition. For instance, as mentioned

above, Palmer [170] showed that context facilitates object recognition, and participants were able to recognize target objects more quickly in a familiar context than in an unfamiliar one. Familiar context sets our expectations about the objects we see in a scene. In return, it facilitates object recognition. Boyce et al. [183] also showed that a contextually consistent background facilitated object perception compared to a meaningless background. Expectations facilitate object recognition by narrowing down the number of possibilities incoming sensory information can give rise to. The most evident examples of the role of top-down influences/expectations come from examples when the object itself is ambiguous, noisy, or incomplete but the context plays an important role in identifying it. Figure 3.1 right panel shows a blurred image of a snowy street with trees around the road. The left column shows the parts of the image alone, representing a cloud, a street light, and a tree, that are ambiguous and unrecognizable. But as soon as we see the image on the right, we can instantly recognize these objects in context based on our expectations about what a road looks like, the presence of a street light at the edge of the road, the presence of clouds in the sky and what a tree looks like on a snowy day half-covered with snow.

2) Expectations help in the visual search of objects in a scene, making it efficient. The above point focuses on the recognition of fixated target objects, but expectations also facilitate the recognition of non-target objects. In Figure 3.2, if you are asked to search for a window, you will look at the buildings and find it. Or if you are asked to search for a car, you will instantly look at the road and find it. However, when you are asked to search for a speed limit sign, you will look around the road and take more time to search for it because it is at an unexpected location. Hence, expectations facilitate visual search. Instead of looking at every element of the scene in depth and identifying objects, our brain uses learned associations/expectations to make the visual search computationally less expensive. Many studies have shown that expectations facilitate visual search [184, 185, 186].

3) Expectations help guide our interactions with objects and actions. Objects are made of or consist of materials that determine their physical properties. Through a lifetime of experiences, we form long-term expectations about the

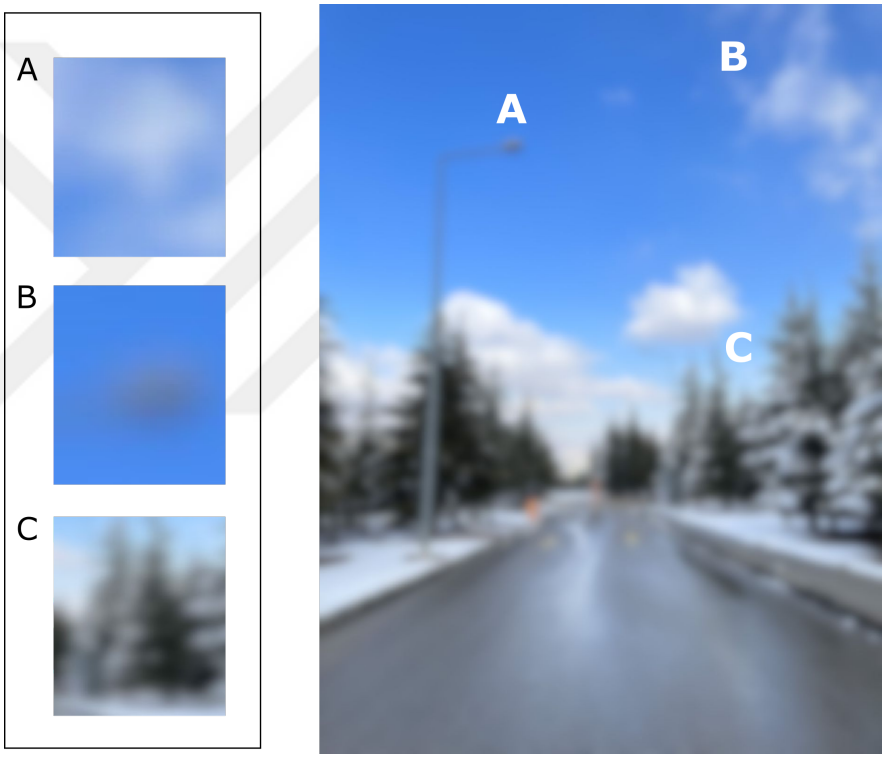

Figure 3.1: An example demonstrating the role of expectations in recognizing objects when the object itself is ambiguous. The left column shows three objects, A, B, and C representing a could, a street light, and a tree which are unrecognizable. The image on the right shows the whole scene from which the left image parts are taken from the locations labeled A, B, and C. All objects are recognizable when put into the context, even though the whole image is blurred.

Figure 3.2: An example demonstrating the role of expectations in visual search in a complicated scene. The left column shows four objects (a cone, a car, a window, and a speed limit sign). The image on the right shows the whole scene from which the left image parts are taken. Finding the expected objects (a cone, a car, or a window) is easy. But the speed limit sign is at an unexpected location (look at the top left corner of the wall of a building). Therefore, it takes longer to find it.

associations between objects and their physical properties [187, 188]. Based on these learned associations, we can predict future states of objects under different forces [189]. For instance, when we hold a teacup in our hand, we will be careful not to drop it because we can predict what will happen if it falls to the ground. On the other hand, we would not worry a lot if a piece of cloth slipped from the grip of our hand. Thus, expectations help us to interact with objects in our daily life efficiently [189, 190, 191].

4) Expectations affect perception by modulating perceptual performance. Research has shown that expectations alter the participant's performance in a perceptual task or affect the speed of perceptual decisions. Downing [192] investigated the perceptual sensitivity of stimuli at different spatial locations for different tasks. He found that expectations, in addition to attention, affected perceptual sensitivity. He showed that perceptual sensitivity is greater when participants were shown an expected stimulus in the attended visual field than the unexpected stimulus in the attended visual field. Expectations have also been shown to speed up perceptual decisions, i.e., expected stimuli are processed faster than unexpected stimuli [193]. Recently, Urgen and Boyaci [194] found that the temporal threshold in an individuation task is increased when expectations are not met compared to a neutral condition, but it remains unchanged in the expected condition compared to a neutral condition. They argued that expectations do not speed up perceptual decisions. Rather sensory processing is delayed when expectations are not met. Therefore, unexpected stimuli take more time to be processed. Nevertheless, it has been well-established that expectations do influence the perceptual sensitivity and speed of perceptual decisions [193, 195, 196, 197, 198, 199].

5) Expectations and predictions ease down the computational burden by not attending to the irrelevant objects. Processing the complex incoming sensory information in the real world can be computationally expensive and cognitively demanding. Expectations help guide our attention to behaviorally relevant stimuli and reduce the computational burden. For instance, if you are driving a car in a crowded area, you will automatically pay more attention to intersections and crosswalks. From your prior experiences, you expect

behaviorally relevant events such as stopping or slowing down the car to occur at these locations. It is inefficient for our visual system to allocate resources to the events/objects that are irrelevant to behavior. For instance, attending to the buildings on the side of roads will be mostly useless to a driver.

6) Expectations can bias perception when the incoming sensory information is ambiguous. In a scenario when incoming sensory information is very weak and ambiguous, the effects of expectations can be very dramatic. For instance, Chalk et al. [200] have shown that expectations bias the perceived direction of motion and can even induce visual hallucinations in the absence of a stimulus. Similarly, Sterzer and colleagues [201] found that expectations can bias the perception of moving dots in ambiguous directions to either moving leftwards or rightwards. One of the classic examples is 'light from above prior', where the perception of convex or concave shape is influenced by our expectation that light generally comes from the above [202]. Another classic example is 'slow speed prior'. According to Weiss et al., [203], we live in a world where most of the objects around us are moving slowly or are static. Therefore, we have an expectation for the objects to move slowly. This expectation is so strong that the speed of an even faster-moving object is perceived as slow.

7) Expectations help us to navigate complex scenes and environments by suppressing input based on statistical regularities. Several studies suggest that stimuli that rarely occur in the natural world or are unexpected usually go unnoticed in complex scenes, a phenomenon that has been called inattentional blindness [204, 205]. A classical example was demonstrated by Simons and Chabris [206], showing that a moving gorilla goes unnoticed in a scene by observers engaged in a counting task. Moreover, expectations help to suppress background noise in complex scenes. For instance, if we are in a crowded place and have to locate our friend, we can easily do so based on our expectations about that friend's appearance.

3.1.2 Long-term and short-term expectations

Expectations can be characterized based on their generalizability over context and time into long-term (structural) and short-term (contextual) expectations [207].

Long-term expectations: Expectations learned over lifetime exposure to statistical regularities in the real world are called long-term expectations. 'Light from above prior' as shown in Figure 3.3 is an example of long-term expectations which has been well established in both human models and animals [208, 209, 210]. Similarly, convexity prior posits strong structural expectations in the figure-ground separation task. Most objects in the natural environment are convex in shape. Hence, observers tend to perceive convex shapes as foreground [211, 212].

One of the primary characteristics of Structural expectations is that they are hard-wired. Sometimes these expectations can also be innate. For instance, many studies show that infants also achieve figure-ground separation using the hard-wired priors described above [213, 214, 215, 216]. However, not all long-term expectations are necessarily innate [207].

Short-term expectations: Expectations developed over a shorter period of time in a particular spatial and temporal context are called short-term expectations. For instance, the ambiguous image in Figure 3.4 can be perceived as a duck or a rabbit depending on the context [217]. In the US, during October at Thanksgiving, people tend to perceive this image as a duck. During easter time, people tend to see it as rabbit since easter is associated with rabbit [217]. Similarly, seeing an image of ducks beside this ambiguous picture will make you perceive it as a duck, while seeing a picture of a bunny can make you see a rabbit [207]. These examples demonstrate that short-term expectations can be formed based on spatial and temporal contexts. However, these expectations are specific to the context they are learned over and do not transfer to other contexts. For example, a study demonstrated that by using the color cue, participants were able to develop expectations that the source of light is not always above but can be at different locations in a given experimental context. But these expectations

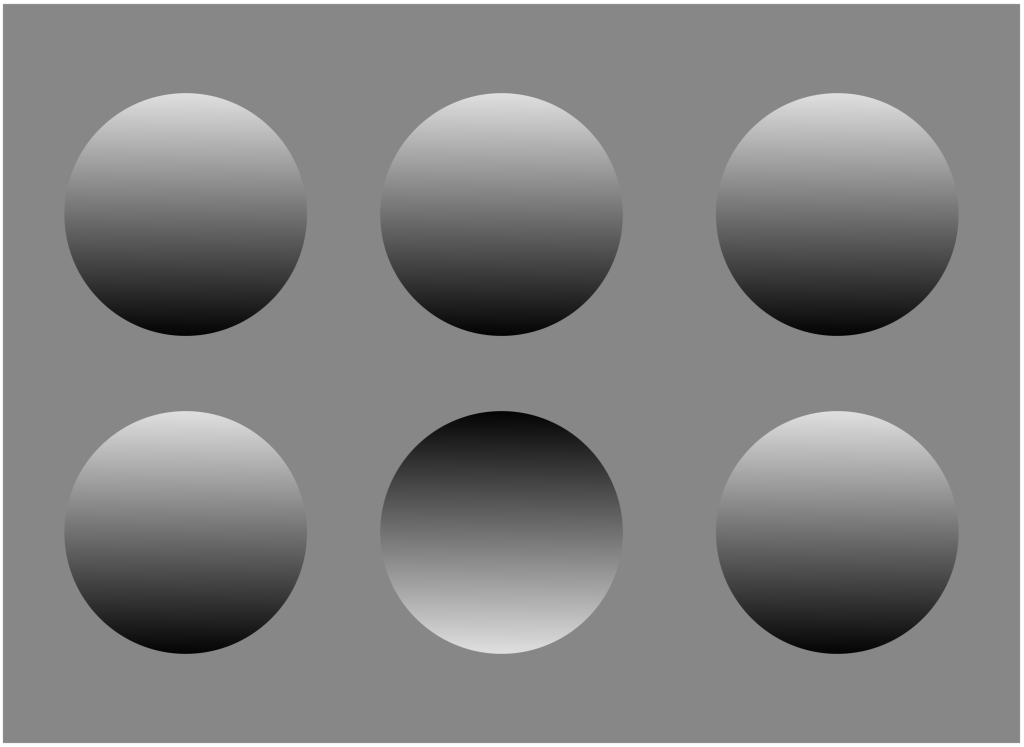

Figure 3.3: An example demonstrating 'Light from above prior' as a long-term expectation. All shapes except the one in the middle of the bottom row are perceived as convex. This is due to our expectations from the real world where light (illumination source) generally illuminates from above.

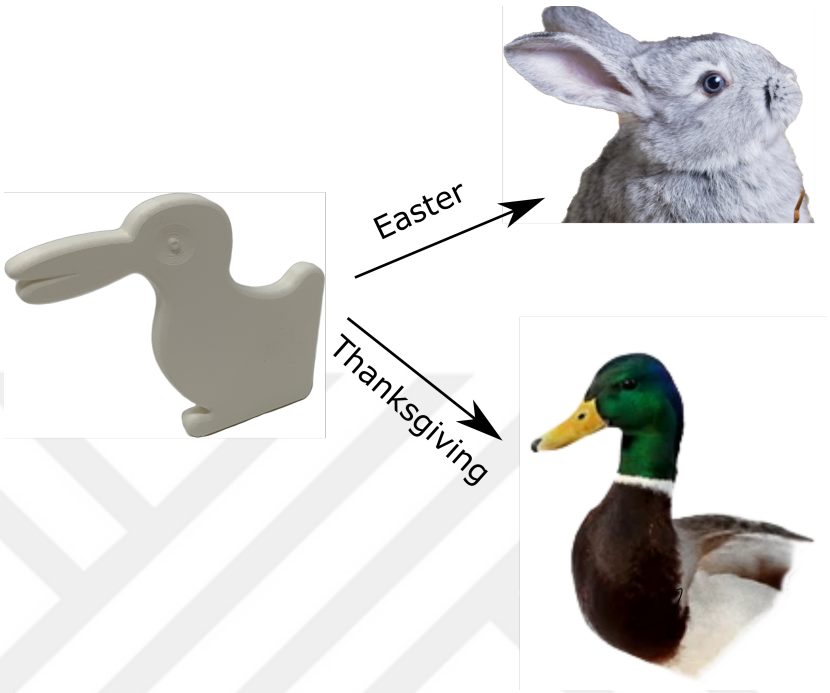

Figure 3.4: An example demonstrating rabbit duck illusion as contextual (short-term expectations). The same ambiguous object is more likely to be perceived as a bunny at easter time and a duck during Thanksgiving time.

were specific to the experimental context. As soon as observers go back to the natural world, light from above prior comes into play again [218].

Can new expectations be learned? While the sensory information is being processed by the brain, it stores information about statistical regularities in the world, including frequency of occurrence of stimulus, conditional probability, and temporal autocorrelations. These statistical regularities lead to the development of expectations to predict incoming sensory information in the future [193]. Hence it is very reasonable to assume that long-term priors actually represent the natural world statistics. Studies have tried to recover long-term priors humans use and showed that at least some of these priors match the statistics of the natural environment [219, 220]. Girshick et al. [219] investigated participants' performance in an orientation estimation task and showed that performance was the best for cardinal orientation, and there was a strong bias to cardinal axes in the presence of uncertainty. They also investigated the distribution of orientations in a set of photographs and found that distribution was dominated by cardinal orientations.

Hence, they showed that the bias is actually reflected in the natural distribution of orientations in the world. Nevertheless, long-term priors, even some hard-wired ones, can be modulated by experimental manipulation. For instance, Knill [221] investigated the effect of short-term expectations on the perception of depth. He found that participants were biased to perceive elliptical shapes as circles presented at an oblique angle. With the training, they were able to eliminate this bias. Moreover, participants gave more weight to the stereoscopic cues rather than shape cues to estimate the slant of the stimulus. Chalk and colleagues [200] also suggest that new priors can be learned in a motion direction task. However, newly acquired priors in experimental settings are short-term. Moreover, some long-term expectations can be easily overridden while others take more time [207]. What is the interplay between these short-term and long-term expectations? Why can some long-term expectations be overridden quickly while others are not? These are some of the questions that need to be investigated [207].

3.1.3 Computational basis of expectations - Bayesian Model

Despite extensive research on the behavioral effects of expectations, computational mechanisms our brain uses to combine incoming sensory information with prior information and expectations are not entirely understood [222]. One of the commonly used models is the Bayesian model, which can account for the effects of expectations. The Bayesian model combines incoming sensory information with prior information and provides an optimal estimate according to Bayes' rule [223]

$$p(w|obs) = \frac{p(obs|w)p(w)}{p(obs)}, \tag{3.1}$$

where w is the world state and obs is the observation/stimulus. According to Bayes' rule, the posterior probability, $p(w|obs)$, of a world state w given the observation obs is the product of the likelihood of observation, $p(obs|w)$, and the prior probability of the world state, $p(w)$, divided by a normalization constant, $p(obs)$ so that the sum of probabilities of all worlds states is equal to one. Figure

3.5 shows the layout of the steps of modeling the behavior under the Bayesian framework [224], which is explained below.

Generative Model: The first step is specifying the generative model. A generative model is the description of the process by which the data/observations (obs) are generated from the world state (w). It consists of the distributions specifying the probability of world state $p(w)$ and the probability of observation given the world state $p(obs|w)$ as shown in the top blue box of Figure 3.5. Expectations in the Bayesian framework come into play at this stage in the form of the probability distribution of the world state $p(w)$ and the probability of observation given the world state $p(obs|w)$. $p(w)$ specifies the statistical regularities in the world. In other words, how often each world state w occurs in the world. For the sake of convenience, this is often taken as a normal distribution. But it can have other distributions as well. An expected world state here will have a higher probability than the unexpected one. The second probability distribution $p(obs|w)$ in the generative model is called the measurement distribution. It specifies that for a given world state, what is the probability of a range of measurements? In simple words, what is the probability that we will get a particular measurement obs when the same world state w is encountered again and again? So, it is centered around the world state of interest w. The variance of measurement distribution is σ^2, which is called measurement noise. If a stimulus is ambiguous, it has higher noise and higher σ^2.

Inference: The observations we make are noisy and incomplete. Under the Bayesian model, we make a probabilistic estimate of the world state using current observation and prior knowledge. Through a lifetime of experience, we already have an expectation about the world states. This expectation is reflected in the natural world statistics, i.e., the stimulus distribution $p(w)$. So, the stimulus distribution often acts as a prior in the inferential process. According to the Bayesian model, we infer the world states by multiplying this prior with the likelihood function as shown in the middle yellow box of Figure 3.5. The likelihood function is derived from the measurement distribution $p(obs|w)$. The difference between measurement distribution and likelihood is that likelihood is a function

of the world state but not the observation. Both prior and likelihood are subjective, while measurement and stimulus distributions are objective. The σ of the measurement distribution specifies the spread of measurement. However, the σ of the likelihood distribution is interpreted as the observer's uncertainty. The output of Baye's rule gives the posterior distribution of the hypothesized world state, $p(w|obs)$, given that particular observation from likelihood and prior. As shown in the distributions on the right side of the yellow box, when the likelihood function is narrower, i.e., the uncertainty associated with the observation is less, the posterior is more towards the likelihood. Otherwise, when the uncertainty associated with the observation is higher, the posterior is more drawn towards the prior. In other words, in the case of noisy and ambiguous incoming sensory information, Bayesian model predicts a bias toward the most probable state of the world based on prior knowledge or expectations.

After the estimation of the posterior probabilities of different world states, A decision is to be made to take action. Under the Bayesian decision theory, a cost function calculates the cost associated with making each decision, and the decision with the highest expected gain is chosen as the final decision. One of the common decision rules is the maximum a posterior (MAP) estimate, in which the observer chooses the world state with the highest posterior probability. Alternatively, the decision may be made based on the posterior mean (PM) as represented by ($\hat{w}$) in Figure 3.5.

Action/Response Probability. The inference process above shows the estimate observer makes on one trial. On repeated presentation, the posterior estimate varies from trial to trial. Thus, the response of the observer over trials itself is a distribution called the response distribution ($p((\hat{w})|w)$) [224]. This distribution helps to compare the behavior of an ideal observer under the Bayesian model to the observer's actual response.

A large body of research literature shows that such a framework could explain how the brain infers the true world states based on noisy and ambiguous incoming information. The formal theory put forward based on such framework is called Bayesian brain hypothesis [1, 221, 225]. Many domains in neuroscience

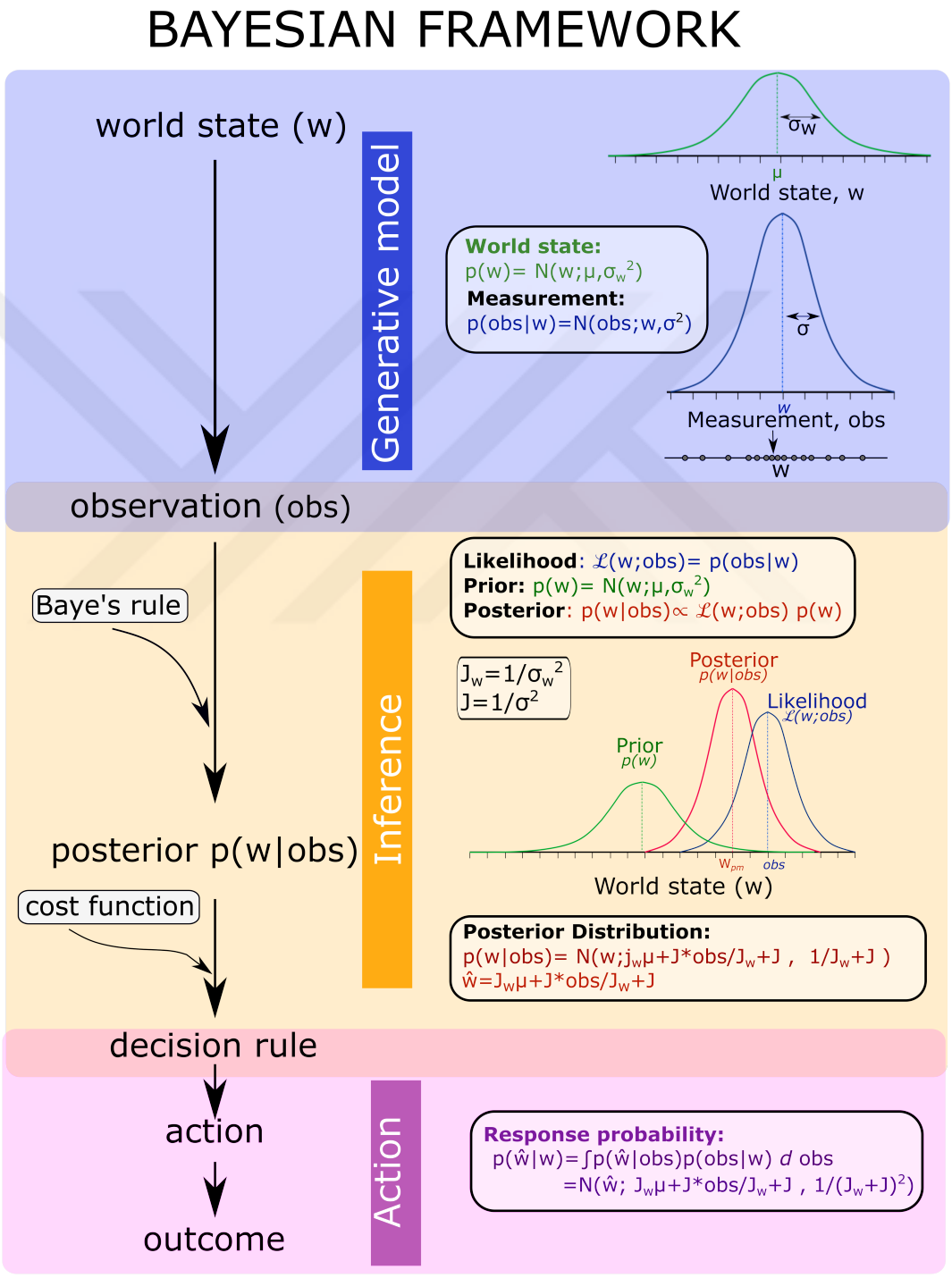

Figure 3.5: Bayesian framework for modeling the effect of expectations on perception. See text for details.

show that human observers perform according to optimal Bayesian inference, for example, detecting the direction of motion [200], planning the movement [226], in contour grouping [227], estimating the speed of a stimulus [203], estimating the orientation of a stimulus [219] or even in everyday cognitive judgments [228]. Such a framework has also been proven to be very effective in modeling the effect of expectations on perceptual decisions [198, 199, 200, 201, 203].

3.1.4 Neural basis of expectations

The Bayesian framework has shown success in accounting for the results of psychophysics studies on expectations. However, how this framework is implemented at the level of neurons and what is the cortical representation of expectations is not fully understood yet [229, 230]. A reasonable neural account for expectation is the modulation of early visual areas by top-down connections from higher areas. Higher regions like parahippocampal and retrosplenial zones have been shown to be associated with object recognition with strong contextual associations. These regions lie more ventral in the frontal lobe and more medial in the posterior cortex in comparison to the regions involved in top-down attentional effects [231]. The orbitofrontal cortex (OFC) is another region of importance in mediating top-down effects in object recognition. OFC provides initial prediction from the association information from low-level areas and feeds back to the extrastriate visual cortex to bias the recognition of contextually familiar objects [232, 233, 234]. Another interesting insight regarding the neural correlates of expectations has been provided by de Lange et al. [235], who measured neuronal responses in the human visual and motor cortex corresponding to predicted, unpredicted, and counter-predictive stimuli for motion direction detection tasks using MEG. They found that participants' decisions could be partly predicted by spontaneous biases in the motor cortex in non-predictive trials showing that perceptual expectations are reflected in the motor system activity and facilitate decision-making. In another interesting study, Fiser et al. [236] proposed that posterior distribution in the absence of incoming sensory information is just a prior distribution. So, the background neuronal activity patterns in the absence

of stimuli (spontaneous activity) just represent prior knowledge that is acquired over the lifetime of experiences. Later, using awake ferrets as animal models, Berkes et al. [237] recorded spontaneous neural activity in V1 at different stages of development. Then, they compared this spontaneous activity with that of the activity elicited by natural images. They found the similarity between the two increases with development. However, it is unclear whether developmental processes or visual experience cause this increase in similarity.

Another important question is how the activity levels of areas reported to be involved in expectations are modulated by top-down effects. Do expectations cause an increase or decrease in neural activity? Evidence of increased neural activity in higher areas corresponding to expected stimuli comes from several electrophysiological studies [238, 239, 240]. For instance, Basso and Wurts [240] recorded single-cell neuronal activity in the superior colliculus of monkeys for a task in which the probability with which the target appeared was manipulated. They found that neural activity was enhanced for the stimuli with a higher probability. Platt and Glimcher [238] have also shown an increase in the activity of neurons in the lateral intraparietal area. On the other hand, many studies suggest that expected stimuli elicit weaker neural response [168, 241]. Summerfield et al. (2008) suggested that the suppression of neural activity on the repeated presentation of stimuli (repetition suppression) could be associated with expectation suppression [168]. They found that the presentation of expected stimuli elicited less neural activation, which means higher suppression. Later, this finding was replicated by a number of studies [241, 242, 243, 244, 245]. However, one of the studies also reported that this expectation suppression could not be found in natural stimuli, suggesting that it is not generalizable over all the stimuli [246].

The findings from the single-cell recordings and neuroimaging studies are inconsistent. A possible explanation for these contradictory findings can be that at the single neuronal level, the response of neurons that respond to expected stimuli becomes greater while the response of neurons that are specific to unexpected stimuli becomes dampened. So, the overall activity of the neuronal population when measured with fMRI decreases. It has been found that the tuning curve of the cortical neurons to preferred stimuli is sharpened by inhibition which also

supports the possibility described above [247]. Kok et al. (2012) also showed that even though neuronal activity in V1 decreases for predicted stimuli, the representation of stimuli gets better when inspected with MVPA [196]. Another possibility is that it is the unexpected stimuli that require extra processing and hence, elicit a higher neural activity than the expected stimuli [194, 248].

3.2 Material perception and expectations

3.2.1 Living in a material world

Expectations play a crucial role in visual perception influencing various aspects, from object recognition to scene perception. Here, we are specifically interested in material perception. From having tea in the morning to driving to the office, working on laptops, and doing dishes, we keep interacting with objects the whole day, and we are always estimating their material properties either by direct physical interaction or just by looking at an object/material. For instance, we don't hold a hot pan with its hot edge; instead, we hold it with a handle. We will use a cup sleeve at a coffee shop to hold a hot cup of coffee. We use a butter knife to cut soft food items like cake but a sharp knife to cut vegetables. Generally, we have a very good idea of what an object is made of, what it looks like, and what material properties it holds. We are able to make decisions about our actions because of our ability to learn these associations. This helps us to interact efficiently with objects in a dynamic environment. There is a huge amount of literature on object perception. Material perception research has also been expanding enormously over the past two decades, but how our brain combines prior learned association with real-time incoming sensory information about material properties is a question yet to be addressed.

3.2.2 How do we perceive Stuff?

Material perception is the identification and recognition of the materials objects are made of. Material perception does not merely rely on visual input but also on the input from different modalities like touch, auditory input, and olfactory input [249, 250]. However, the main focus of the research had been on the visual aspect of material perception. This is not unexpected given the advancement in computer vision, artificial intelligence, and neural networks contributing to the development of more sophisticated computer vision software and photorealistic images [251]. Another factor is that vision plays the most important part in material recognition. It is impressive how humans are able to identify materials without any effort based on the visual information about surface structures that may appear meaningless, particularly when considering the fact that many different material groups share the same visual characteristics, as shown in Figure 3.6. Not only this, the same objects made of the same material category may have different visual properties of the surface (see Figure 3.7). In spite of these ambiguities, studies have demonstrated that material perception is a very fast and accurate process not relying on the low levels features [252, 253]. On the other hand, there are many studies from the computer vision domain that argue that low-level features are required to categorize materials [254, 255, 256]. However, recognizing materials using only low-level features is difficult in real-world scenarios where we know that many objects having similar low-level properties can be made from different materials. Therefore, it is suggested that prior knowledge plays a very important role in the recognition of material and material properties. Fleming et al. [257] designed two experiments where participants were asked to assess the material properties of different materials using both visual images and verbal names of materials. Their findings revealed a consistent association between the judged material properties and the presented material categories, whereby the judged material properties formed distinctive feature spaces that aligned with the material categories. Additionally, the study identified a link between the two tasks, suggesting that the visual assessment of material properties and the categorization of materials from memory originated from the same stored knowledge of materials. In conclusion, our current understanding suggests that

Figure 3.6: An example illustrating objects made of different materials (plastic and ceramic plaster) but appear similar in terms of their surface appearance.

visual perception of materials is a highly efficient and accurate cognitive process that is distinct from object recognition and is affected by expectations. The effect of expectations on material perception can also be modeled under the Bayesian framework, which will be demonstrated later.

Figure 3.7: An example illustrating the same material category ‘textile’ in four pictures but all are having different visual properties. Despite the differences in low-level visual properties, when observers are asked to categorize them, they will place them all under the ‘textile’ material category.

3.3 Unmet Expectations About Material Properties Delay Perceptual Decisions- Behavioral study.

The studies and results presented in this chapter are based on a publication by A. Malik and H. Boyaci, "Unmet Expectations About Material Properties Delay Perceptual Decisions," which is published in Vision Research and also in BioRxiv.

As discussed earlier, objects are made of or consist of materials that determine their physical properties. Through a lifetime of experiences, we form long-term expectations about the associations between objects and their physical properties [187, 188]. Based on these learned associations, we can predict future states of objects under different forces [189]. Such expectations are believed to influence behavior through top-down processes and may often be implicit [189, 190, 191]. Indeed we become aware of our expectations only when we encounter a situation in which they are unmet or violated, as illustrated in Figure 3.8. Such surprise effects are an important aspect of human experience and are not only of great interest to researchers but also to artists and designers who use them strategically in their works (see, for example, see, for example, Ludden, Schifferstein, and Hekkert 2008, 2009 [258, 259]).

Here we study the effect of long-term and newly acquired, context-dependent expectations about material properties on the speed of relatively low-level perceptual decisions. A great number of studies have shown that observers perceive expected stimuli faster [260, 261, 262]. Those studies, however, usually focused on the identification of static stimuli. Only a few studies tested the effect of expectation about material properties in dynamic scenes [189]. In their study, Alley et al. presented participants with computer animations of objects that were falling down and behaving in a predicted or surprising way upon hitting the ground. For example, a teacup could shatter as predicted or, surprisingly, wrinkle like a piece of cloth. The task of the observers was to judge as quickly and as accurately as possible one of four material attributes of the objects in each trial,

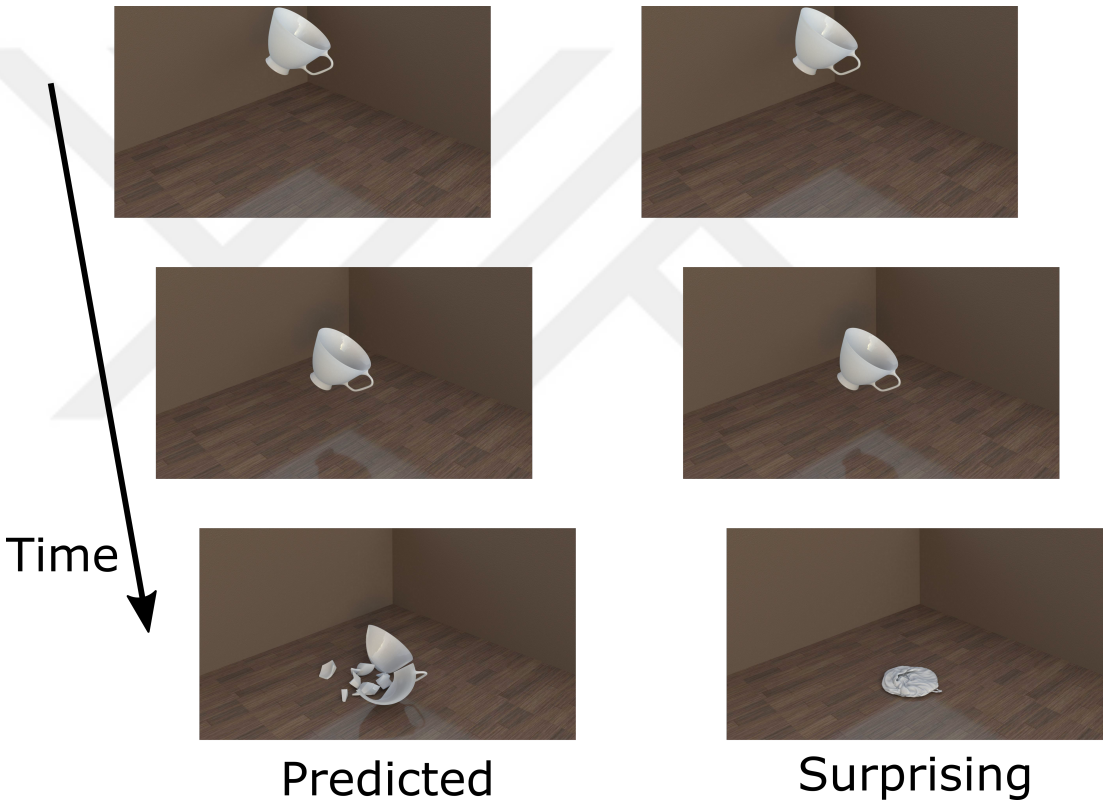

Figure 3.8: As soon as we see a teacup start falling down, our visual system predicts what might happen to it: if not caught, the cup will hit the ground and shatter. If the cup wrinkles as a piece of cloth upon hitting the ground, we will be surprised and even amazed. Because our expectations and the visual input mismatch.

which were hardness, gelatinousness, heaviness, and liquidity. Alley et al. found that long-term expectations bias the perception of material attributes of familiar but not unfamiliar objects. For example, a spoon that wrinkles upon impact was judged harder than a piece of cloth that also wrinkles. Further, they showed that the reaction times were longer in the surprising trials than in expected trials for familiar, but not for unfamiliar objects. In the current study, we use a paradigm similar to that of Alley et al.. We present observers with computer animations of familiar objects falling down and behaving in a predicted or surprising way upon hitting the ground. Unlike in the experiment by Alley et al., we do not ask observers to judge material qualities which engage mid to high-level visual processes [263, 264, 265], but instead target relatively low-level perceptual decisions: the task of the observers was simply to judge "did the object break?" Breaking, after hitting the ground, is usually characterized by a typical motion pattern where a uniform vertical trajectory is followed by a fast radial motion pattern. Thus breaking entails a fairly low-level motion cue, i.e., sudden changes in motion direction and speed that can be detected by low-level visual mechanisms [266]. Importantly, in our experiments, breaking objects always break, and non-breaking objects always remain intact upon hitting the ground in both predicted and surprising conditions. Thus, the correct response for the same object, whether surprising or predicted, does not change, eliminating a response preparation confound. With this paradigm and by assessing reaction times (RTs) of observers' judgments about whether breaking occurred, we are able to assess the effect of expectations about material properties on relatively low-level perceptual decisions.

Our research question is whether expectations about material properties affect low-level perceptual decisions. We hypothesize that if they do, then RTs should be different under the predicted and surprising conditions. To anticipate, under two different experimental manipulations and with two groups of participants, we found that RTs are indeed longer for the surprising trials. Further, we found an interesting interplay between long-term expectations and newly formed expectations based on context-dependent regularities, for which we propose possible explanations.

3.3.1 Materials and Methods

3.3.1.1 Participants

Twenty-eight participants participated in the experiment. All had normal or corrected to normal vision and were naive to the purposes of the experiment. Participants gave their written informed consent before the first experimental session, in line with the guidelines of the Declaration of Helsinki. Experimental protocols and procedures were approved by the Research Ethics Committee of Bilkent University, Turkey.

3.3.1.2 Stimuli presentation

An LCD color reference monitor (Eizo CG2730, 27 inches, 2560 x 1440 resolution, 60 Hz refresh rate) was used for stimulus presentation. The monitor was the only light source in an otherwise completely dark room where the experiment took place. Participants sat on a chair and viewed the monitor from a distance of 60 cm. A chin rest was used to minimize the head movements. The experimental paradigm was programmed with Psychtoolbox [267] on MATLAB version 2018a (MathWorks Inc., New York, NY, USA).

Stimuli were generated by a professional graphic artist (Aleksa Radakovic) using commercial software (Cinema 4d) and consisted of computer animations of six objects that act in a certain way when dropped on the ground; three of them break upon hitting the ground (breaking objects: wine glass, pot, and teacup) and the other three do not break (non-breaking objects: spoon, key, and rod). Each animation consisted of 46 frames. In all animations, the objects hit the ground in frame number 15. There were two animations for each object. In one set of animations, objects behaved in a natural way upon hitting the ground. Specifically, breaking objects shattered, and non-breaking objects bounced rigidly after they hit the ground. In the other set, objects behaved in an atypical way upon hitting the ground: breaking objects graveled and non-breaking objects

wobbled. Figure 3.9 shows examples of these natural and atypical behaviors.

3.3.1.3 Experimental Design

Participants were divided into two groups; natural group and atypical group. Each participant underwent a training session followed by an experimental session. During the training session, participants in the natural group were presented with animations where objects behaved naturally, whereas participants in the atypical group were presented with animations where objects behaved atypically (20 trials for each object). Thus, the newly formed context-dependent expectations about an object's material properties in the atypical group were different from the already existing long-term expectations.

During the experimental session, 10 animations were shown for each object. Of those 10, 8 were the same as in the training session (naturally behaving objects for the natural group, atypically behaving objects for the atypical group). We call these predicted trials. The remaining 2 trials were from the untrained category (atypically behaving objects for the natural group, naturally behaving objects for the atypical group). We call these surprising trials. The order of presentation was randomized in all sessions. Fig. 3.10 shows example sequences of trials for training and typical testing sessions for both groups (natural & atypical).

All sessions started with an instruction screen, followed by the animations as soon as any key was pressed. All animations were preceded by a 1-second blank screen with a central fixation cross. Each animation was 1.53 seconds long (46 frames, 30 frames per second). The task was to answer the question, "Did the object break?" by pressing the corresponding keys for "yes" and "no" on the keyboard after the object hit the ground. In the training session, an error sound was delivered if the participant answered the question before the object hit the ground. Reaction times were measured from the time an object made an impact on the ground (15th frame) to the time when the participant pressed a key. The next trial did not start until the observer responded.

Figure 3.9: Six objects used as stimuli and their natural and atypical behaviors. For the participants in the 'natural group,' natural behavior was predicted, and atypical behavior was surprising. For the participants in the 'atypical group,' atypical behavior was predicted, and natural behavior was surprising. These expectations were formed through a training session before the main experiment. See text and Figure 3.10 for details about the training procedure.

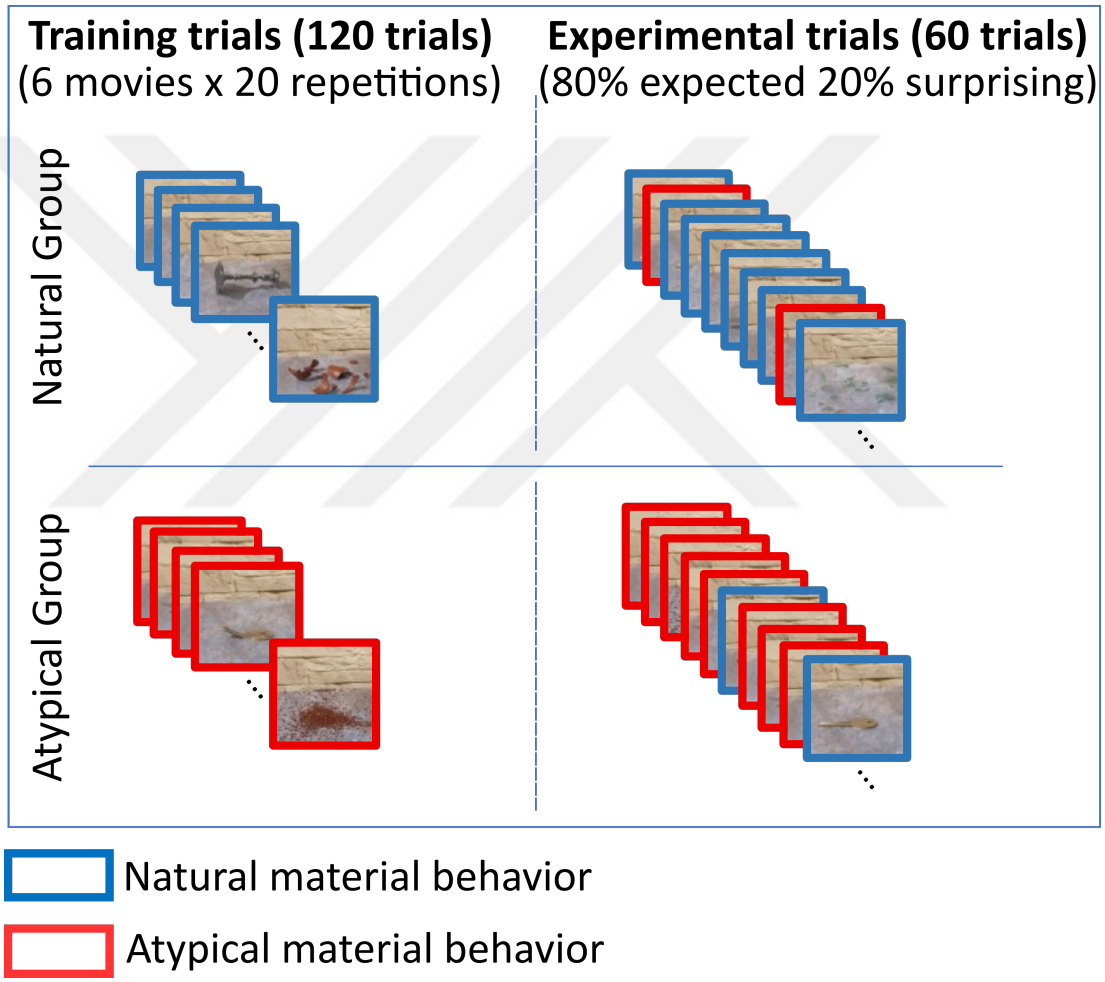

Figure 3.10: Experimental Design: Each participant underwent a training session before the experimental session. Participants in the natural group were trained on naturally behaving objects, while participants in the atypical group were trained on atypically behaving objects. In the experimental session, contextual expectations formed in training sessions were unmet in 20% of the trials.

3.3.1.4 Analysis

Analyses were performed on MATLAB version 2018a (MathWorks Inc., New York, NY, USA) and RStudio [133]. For the training session, data from trials in which reaction times were negative (i.e., a response was made before the object hit the ground) were excluded from analyses. For the experimental session, data from trials in which reaction times were negative or did not fall within the ±3SD of the mean were excluded from analyses (a total of 24 out of 1680 data points excluded).

The histogram of the raw reaction time data indicated that they were skewed and not normally distributed. To remedy this, before any subsequent analysis, we transformed the RTs using a Box-Cox transformation. Note, to make interpretation of the results easier, we use raw RTs in the figures and when we report means in the text. Next, we ran two omnibus Linear Mixed Models (LMMs), one for the training RTs, and one for the experimental RTs, using the 'lme4' package [132] in RStudio [133]. For both models, the transformed reaction times were included as the dependent variable, and the between-subject variance was estimated using a random intercept in the model. For the training, we wanted to investigate to what extent the training modulated reaction times in natural and atypical groups. To answer this question, we wanted to contrast RTs in the beginning (first five trials) and the end of the training session (last five trials) for each group (we call this predictor 'time'). Thus our 'training' LMM model included time, group, and their interaction as fixed effects:

$$RT \sim Time \times Group + (1|\, participant\ ID). \tag{3.2}$$

For the experiment, we wanted to find out to what extent RTs might be modulated by the experimental condition (predicted vs. surprising), the type of the object, i.e., how its material reacted to impact force (breaking vs. non-breaking), and the group (natural vs. atypical). Thus the 'experiment' model included condition, group, object, and their two-way and three-way interactions as fixed effects:

$$RT \sim Condition \times Group \times Object + (1|\, participant\ ID). \tag{3.3}$$

To calculate the significance of the fixed effects, we used the 'lmerTest' package [134], which uses Satterthwaite's method to estimate the degrees of freedom and generate p-values for linear mixed models. Any potential post hoc analysis for pairwise comparisons of estimated marginal means was done with the 'emmeans' package [135] using a Tukey correction.

Percent correct responses were compared between expected and surprising conditions for each group using a two-tailed Student's t-test, correcting for multiple tests (alpha=0.025/2).

3.3.2 Results

3.3.2.1 Training Session

Using linear mixed model analysis that included time (first vs. last five time points) and group (natural vs. atypical) and their interaction as fixed effects, we found that time ($F(1) = 63.3, p < 0.001$), and the interaction of time and group ($F(1) = 19.06, p < 0.001$) significantly modulated RT. The main effect of group was not significant. To follow up the interaction, we conducted pairwise comparisons of RTs within each group and found that only for the atypical group, the difference in RT between initial and final trials was significant ($p < 0.001$). Figure 3.11 shows the mean RTs as a function of trial number in the training session as well as the mean RTs for only the first and last five trials of the training for each group. The latter clearly illustrates the significant interaction found in the LMM, which suggests that training effects were strongest for the atypical group.

3.3.2.2 Experimental Session

Before conducting the LMM analysis on RTs, we wanted to verify that there was no difference between expected and surprising conditions in terms of performance

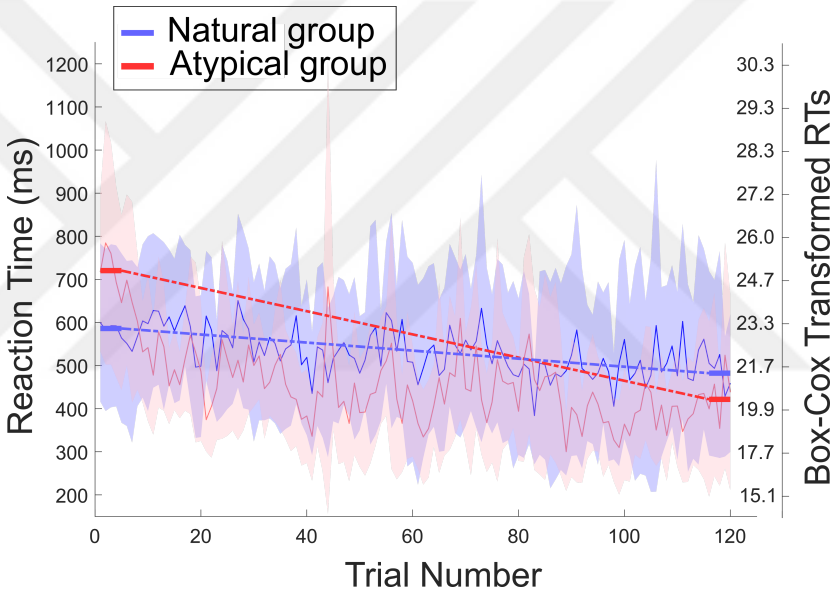

Figure 3.11: Reaction times (RTs) from the training sessions of both groups as a function of trial number. The blue solid line shows mean RTs for the natural group (across 14 participants), and the pink line shows mean RTs for the atypical group (across 14 participants). Shaded regions are the 95 percent confidence interval around the mean. Thick horizontal solid lines show mean RTs on the first and last 5 trials. Overall, RTs get shorter as the session progresses. This training effect is stronger for the participants in the atypical group, who were trained on the atypical behavior.

(percent correct) in the behavioral task (indicating whether an object broke or not). A two-tailed Student's t-test yielded no statistically significant differences between the percentage of correct responses of expected (average 98.6%) and surprising (average 98.5%) conditions. Next, using linear mixed model analysis that included condition (predicted vs. surprising), group (natural vs. atypical), object (breaking vs. not breaking), and their interaction as fixed effects, we found that condition ($F(1) = 60.44, p < .001$), object ($F(1) = 16.54, p < .001$), the interactions between object and group ($F(1) = 29.95, p < .001$), and object and condition ($F(1) = 15.66, p < .001$), as well as the interaction between condition, group, and object ($F(1) = 17.81, p < .001$), all significantly modulated RT. Table 3.1 reports the corresponding estimates of all fixed effects.

Table 3.1: Linear mixed model fixed effect estimates from the experimental session. Estimates and standard errors correspond to Box-Cox transformed RTs. (Significant effects are in bold.)

Predictors	**Estimates**	**CI**	**p-Value**
Intercept	26.04	24.47 – 27.60	**<0.001**
Condition (Predicted)	-0.91	-1.13 – -0.68	**<0.001**
Object (Breaking)	-0.47	-0.70 – -0.25	**<0.001**
Group (Natural)	0.76	-0.81 – 2.33	0.342
Condition (Predicted) x Object (Breaking)	0.46	0.23 – 0.69	**<0.001**
Condition (Predicted) x Group (Natural)	-0.22	-0.45 – 0.00	0.054
Object (Breaking) x Group (Natural)	-0.64	-0.87 – -0.41	**<0.001**
Condition (Predicted) x Object (Breaking) x Group (Natural)	0.49	0.26 – 0.72	**<0.001**

Figures 3.12-3.14 aid in interpreting the significant effects found in the LMM analysis. Overall, across conditions, RTs were longer in the surprise condition (meanRT= 619.2 ms) than in the predicted one (meanRT= 510 ms; Figure 3.12, difference between dark and light blue bars). The same pattern is present in nearly every participant (Figure 3.13) and is also manifested in the negative value of the estimate of the predictor condition (Table 3.1): the condition 'predicted' is a negative predictor of RT surprising (which is used as the reference category in the model).

The presence of this significant main effect, in essence, would answer our main

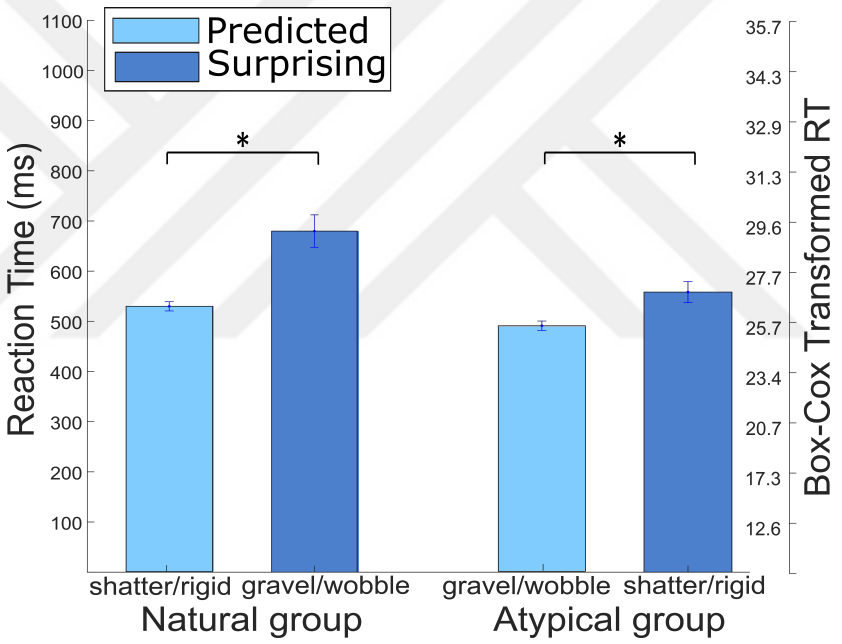

Figure 3.12: Mean RTs of predicted and surprising conditions averaged across participants. When expectations are unmet, whether natural or atypical, perceptual decisions are delayed (post hoc pairwise comparison of estimated marginal means are significant, *: $p < 0.01$). This effect tends to be stronger in the natural group. The error bars represent SEM.

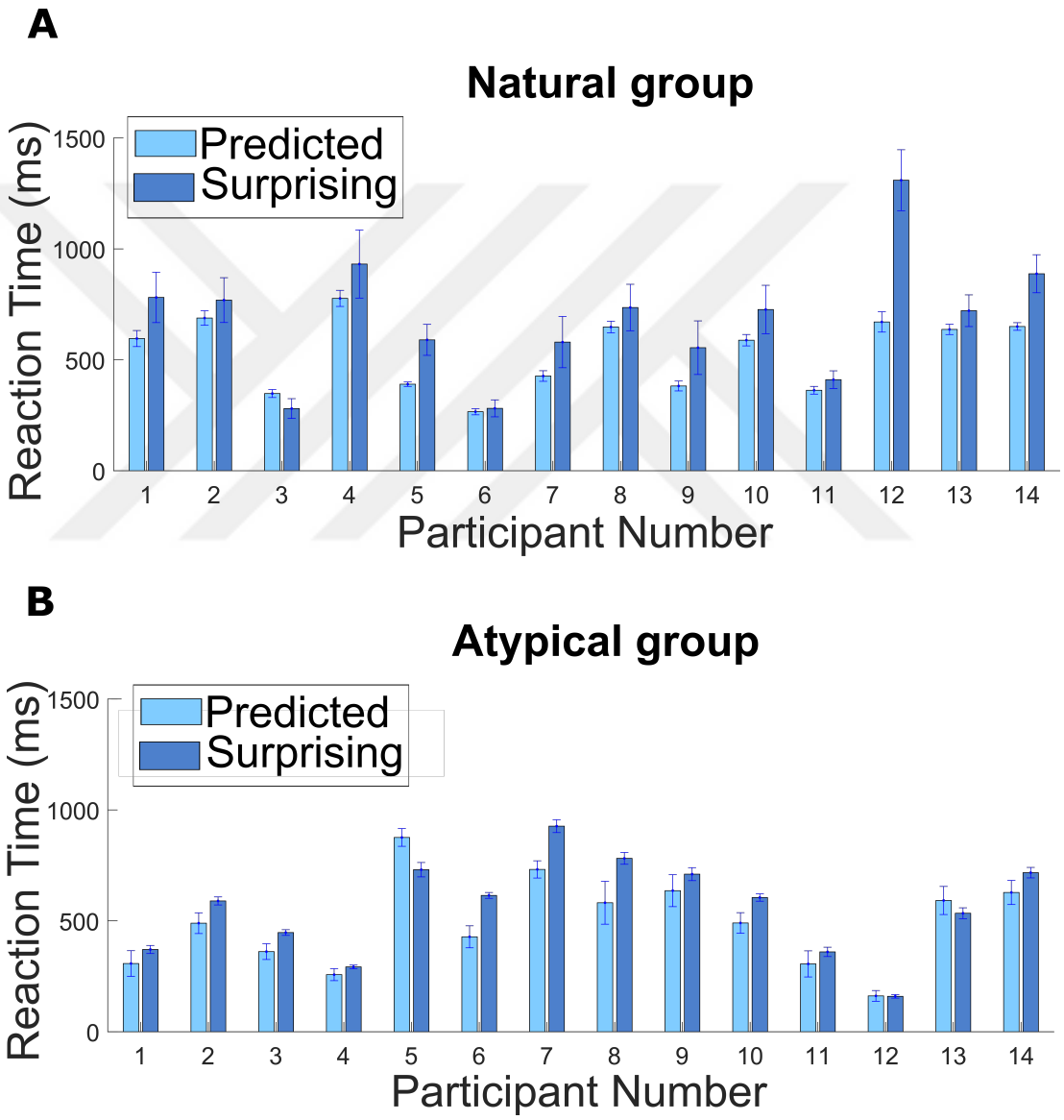

Figure 3.13: Mean RTs per participant in the experimental session for (A) natural and (B) atypical groups. Light and dark blue bars denote mean RTs in the predicted and surprising conditions, respectively. Error bars correspond to SEM.

research question, namely whether expectations affect the speed of making perceptual decisions and whether participants take longer when their expectations are unmet. The main effect of condition, however, has to be interpreted in light of the two- and three-way interaction(s) with object and group, which we will focus on next.

The interaction between condition and group can be understood by inspecting Figure 3.12: while the difference between predicted and surprise conditions was significant for both groups ($p < 0.01$), it was somewhat larger in the natural than in the atypical group. We also found an interaction between condition and object, with RTs under the surprising condition for the non-breaking objects (mean=671 ms) being, on average, significantly longer than those for the breaking objects (mean=576.5 ms, $p < .001$), whereas RTs under the predicted condition did not differ for the two object types (mean for breaking=510.1 ms, mean for non-breaking=509.9 ms, $p = 0.99$). Further, Figure 3.14 reveals the cause for the significant three-way interaction between condition, group, and object: the difference between the RTs of predicted and surprising conditions was significant only for the non-breaking objects in the natural group ($p < 0.001$), and conversely, only for the breaking objects in the atypical group ($p < 0.05$). All significant pairwise comparisons are listed in Table 3.2.

Taken together, these interactions do not invalidate the main effect of condition; instead, they paint a more interesting picture of how expectations affect the speed of making perceptual decisions, which we will discuss below.

3.3.3 Discussion

Here we studied the effect of expectations about material properties on the speed of relatively low-level perceptual decisions. We presented computer animations of objects falling down and asked the participants to report as soon as possible whether the objects broke or not upon hitting the ground. We found that participants were slower to make this judgment when their expectations about the material properties were not met. Furthermore, this was true even when

Table 3.2: Follow-up contrasts for the three-way interaction of condition, group, and object that we found in the experiment. Significant comparisons are marked with bold p value text. PRE: predicted, SUR: surprising, NAT: natural group, ATP: atypical group, BRK: breaking object, NBR: non-breaking object.

contrast	**estimate**	**SE**	**df**	**t.ratio**	**p.value**
PRE BRK NAT - SUR BRK NAT	-0.3549	0.4678	1622.02	-0.759	0.9950
PRE BRK NAT - PRE NBR NAT	-0.3175	0.2945	1622.00	-1.078	0.9613
PRE BRK NAT - SUR NBR NAT	-4.4832	0.4632	1622.00	-9.679	**<.0001**
PRE BRK NAT - PRE BRK ATP	0.7785	1.6081	27.07	0.484	0.9996
PRE BRK NAT - SUR BRK ATP	-0.6455	1.6480	29.86	-0.392	0.9999
PRE BRK NAT - PRE NBR ATP	1.0447	1.6082	27.09	0.650	0.9977
PRE BRK NAT - SUR NBR ATP	-0.2568	1.6486	29.91	-0.156	1.0000
SUR BRK NAT - PRE NBR NAT	0.0374	0.4677	1622.03	0.080	1.0000
SUR BRK NAT - SUR NBR NAT	-4.1282	0.5887	1622.02	-7.013	**<.0001**
SUR BRK NAT - PRE BRK ATP	1.1334	1.6486	29.91	0.688	0.9967
SUR BRK NAT - SUR BRK ATP	-0.2905	1.6876	32.83	-0.172	1.0000
SUR BRK NAT - PRE NBR ATP	1.3996	1.6487	29.92	0.849	0.9884
SUR BRK NAT - SUR NBR ATP	0.0981	1.6882	32.88	0.058	1.0000
PRE NBR NAT - SUR NBR NAT	-4.1656	0.4630	1622.00	-8.996	**<.0001**
PRE NBR NAT - PRE BRK ATP	1.0960	1.6080	27.07	0.682	0.9969
PRE NBR NAT - SUR BRK ATP	-0.3280	1.6479	29.86	-0.199	1.0000
PRE NBR NAT - PRE NBR ATP	1.3622	1.6082	27.08	0.847	0.9884
PRE NBR NAT - SUR NBR ATP	0.0607	1.6486	29.91	0.037	1.0000
SUR NBR NAT - PRE BRK ATP	5.2616	1.6473	29.81	3.194	0.0574
SUR NBR NAT - SUR BRK ATP	3.8377	1.6863	32.73	2.276	0.3364
SUR NBR NAT - PRE NBR ATP	5.5278	1.6474	29.82	3.355	**0.0397**
SUR NBR NAT - SUR NBR ATP	4.2264	1.6869	32.78	2.505	0.2286
PRE BRK ATP - SUR BRK ATP	-1.4240	0.4651	1622.00	-3.061	**0.0463**
PRE BRK ATP - PRE NBR ATP	0.2662	0.2948	1622.02	0.903	0.9858
PRE BRK ATP - SUR NBR ATP	-1.0353	0.4674	1622.01	-2.215	0.3432
SUR BRK ATP - PRE NBR ATP	1.6902	0.4657	1622.01	3.629	**0.0071**
SUR BRK ATP - SUR NBR ATP	0.3887	0.5903	1622.00	0.658	0.9980
PRE NBR ATP - SUR NBR ATP	-1.3015	0.4680	1622.02	-2.781	0.1005

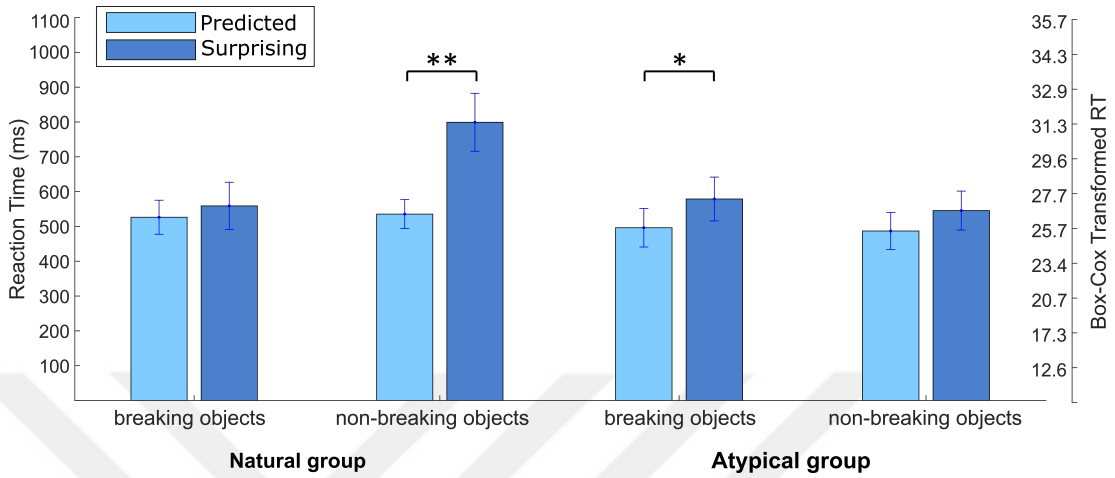

Figure 3.14: Mean RTs in the experimental session for breaking and non-breaking objects. The left side shows the results of the natural group, while the right side shows the results of the atypical group. Posthoc pairwise comparison of estimated marginal means is marked significant with ** for $p < 0.001$ and * for $p < 0.05$. Error bars represent SEM. For all significant pairwise comparisons, see Table 3.2.

participants were trained to predict an atypical behavior, for example, that a candle stick wobbled as if made of jelly. The pattern of our results can not be explained by motor response preparations because whether under the predicted or surprising condition, for a given object, the correct response was always the same: breaking objects always broke, and intact objects always remained intact. Motion statistics, on the other hand, might have affected the RTs. For example, it could be easier to decide that an object remains intact with the motion statistics of a rigid body compared to a gelatinous one. Those low-level motion statistics alone, however, cannot completely explain the differences between the two experimental groups. Because if motion statistics alone determined the responses, we would find the same pattern of RTs for both groups; namely, participants would always be quicker on the trials of a certain motion type, for example, on the trials of rigid body motion.

3.3.3.1 Expectations about material properties affect low-level perceptual processes.

Our main finding is in line with previous studies. For example, [189] found that unmet expectations delay participants' decisions about material attributes. But unlike in most previous literature, in our study, the participant's task was not about material attributes. And the participants did not need to attend to and process the material properties; they only needed to analyze the motion patterns after the objects hit the ground. Thus, a sensible strategy could have been to ignore or down-weigh the object-material associations and instead focus entirely on the low-level motion patterns after the impact. Yet, we find that participants' decisions were nevertheless affected by their expectations about material properties. This demonstrates that high-level expectations can affect low-level perceptual processes, even when those expectations are task-irrelevant.

3.3.3.2 Training alters expectations.

Our daily subjective experiences and previous research on the topic [189, 268, 265, 269] suggest that we form associations between an object and its typical material properties. These associations not only help us to recognize and identify the object and materials efficiently but also help in action planning and guiding our interaction with them [187, 270, 271]

Some long-term expectations are "stubborn" and do not easily change, but some can be altered under experimental conditions [193, 272]. For example, [208] showed that "light from above" prior could be altered when participants are trained with haptic feedback. Similarly, [273] showed that "slow speed prior", which explains many motion and direction illusions, can be altered through training sessions. The pattern of RTs we found in the current study is consistent with this literature. We found that RTs of the atypical group were longer under the surprising condition compared to the predicted condition, even though the predicted atypical behaviors were in conflict with the long-term expectations.

In essence, participants learned new context-dependent expectations during the training session.

RT data from the training session provides further insights into the progress of this learning. Firstly, the decrease in RTs was larger for the atypical group compared to the natural group. This was anticipated because only in the atypical group did participants learn new associations and form new context-dependent expectations. At the beginning of the training sessions, RTs of the atypical group were longer than those of the natural group, which was also anticipated because the object behaviors were atypical and not predicted based on long-term expectations. But as the session progressed, the atypical group participants started to learn to expect the atypical behaviors in the context of the experiment, and their RTs decreased. Towards the end of the session RTs of the atypical group were equal to, and even slightly lower than RTs of the natural group. This further reduction might be related to an 'oops' factor, whereby a sequence of asynchronously presented mismatching cues can lead to efficient learning [274].

3.3.3.3 Interplay between long-term expectations and context-dependent regularities.

The overall effect of expectations, i.e., a larger RT difference between predicted and surprising conditions, tended to be stronger in the natural group compared to the atypical group. For the natural group, where long-term expectations and context-dependent learned regularities were consistent, a strong expectation effect was indeed anticipated. Whereas for the atypical group, long-term expectations, which can often be strong [207], were in conflict with the newly acquired expectations. This conflict could have reduced the overall strength of the newly acquired context-dependent expectations in the atypical group. Further scrutiny revealed a significant effect of expectation for intact objects but not for breaking objects in the natural group. Conversely, for the atypical group, there was a significant effect for breaking objects but not intact objects. This finding might seem puzzling at first, but it can be explained by different strengths of long-term expectations. Long-term expectations for the non-breaking objects used in the

experiment, such as the candlestick being rigid - rather than gelatinous, might be very strong, leading to the significant effect found for those objects in the natural group. These long-term expectations, however, strongly conflict with the context-dependent regularities for the atypical group and thus produce weaker new expectations and result in no effect for the non-breaking objects in that group. Conversely, for the breaking objects used in the experiment, long-term expectations to shatter might not be much stronger than the long-term expectations to gravel, leading to little or no effect of expectation in the natural group. But this time, because the long-term expectations are weak, the newly-acquired expectations are stronger, and this results in a significant effect for the atypical group. Essentially, the pattern of our results indirectly reveals novel facts about the strength of various expectations about material dynamic properties.

Note that, in principle, these conceptual arguments can be formalized into a computational framework based on the Bayesian model explained in the introduction section. Traditionally, in a Bayesian model, as explained previously, the posteriors are calculated based on the prior distributions and the observed data. However, our proposed approach involves an iterative process in which posteriors are computed iteratively in time [275, 276]. This iterative calculation takes into account the updated information from each iteration and refines our understanding of the model parameters. Such a framework is discussed further in the following paragraph. Likewise, more mechanistic, iterative predictive processing models could be employed to establish a link between behavior and neuronal activity [248, 277].

3.3.3.4 A Bayesian updating model can explain the results.

In this section, we conceptually introduce a Bayesian updating model that can explain the pattern of our findings more formally. Bayesian models have been successfully applied to a variety of visual perception problems, including those related to the effects of expectations [161, 229, 278, 279, 280]. In its basic form, the Bayesian rule simply combines a prior probability distribution of the world states and the likelihood of those world states under the observed data. Here

we propose a model in which the prior, representing expectations, is not static but continuously updated in time [275]. The basic equation for the posterior probability distribution, i.e., prediction in our case, becomes

$$p_t(B|\mathbf{v}) \propto p_t(B)p_t(\mathbf{v}|B), \tag{3.4}$$

where B represents possible world states (intact-rigid, intact-wobbling, breaking-shattering, breaking-graveling), $\mathbf{v}$ is an abstraction for the visual motion statistics data, $p_t(\mathbf{v}|B)$ is the likelihood of the world states at time point t, $p_t(B)$ is the prior probability distribution at t, which is updated based on the previous posterior and the long-term prior, p_{lt}

$$p_t(B) = f(p_{t-1}(B|\mathbf{v}), p_{lt}), \tag{3.5}$$

e.g. a weighted sum of the terms, where weights determine how "stubborn" the prior is. Note that p_{lt} implicitly depends on the object:

$$p_{lt}(B) = p_{lt}(B|\text{object}). \tag{3.6}$$

For example, for the participants in the natural group, it could be $p_{lt}(\text{intact-rigid}|\text{candle stick}) \approx 1$ and all else ≈ 0.

We propose that within a single trial, posterior estimates are continually updated iteratively using equation 3.4 over time until a decision is reached. A decision is made by the model using a decision criterion e.g. by comparing the ratio of probability of different world states with a decision threshold. Note that, within a single trial, prior ($p_t(B)$) in equation 3.4 is not fixed, but the posterior computed in the previous iteration becomes prior for the next iteration. According to the model, if the stimulus shown is consistent with the long-term prior (expected stimulus), a decision is reached quicker, i.e. less number of iterations are required to reach the decision threshold (scenario 1 in Figure 3.16). However, if the stimulus shown is inconsistent with long-term expectations (surprising stimulus), the posterior computed would not be sharp enough to reach a decision threshold, and therefore, it will continue to update in time until a sharp posterior is obtained and the decision is made (scenario 2 in Figure 3.16). Moreover, we propose that the prior does not remain fixed over trials. It is updated on

every trial based on the previous trial, as shown in equation 3.5. Over repeated trials, the observer learns new contextual or short-term expectations depending on the stimuli, and hence long-term expectations are updated to contextual or short-term expectations. This is discussed in detail below:

Training Session: When the training starts, participants in both natural and atypical groups have their long-term expectations from prior life experiences i.e. the intact objects behave rigidly while breaking objects shatter. In the atypical group, participants are trained on atypical stimuli i.e., intact-wobbling objects and breaking-graveling objects. Our model predicts that over time with repeated presentation of atypical stimuli, participants learn to associate atypical behavior with objects and develop short-term or contextual expectations under experimental settings, as shown in Figure 3.15. Now, when they are presented with stimuli in the experimental session, they will expect intact objects to wobble and breaking objects to gravel under experimental conditions. Hence, after training, for the atypical group, intact-rigid and breaking-shattering stimuli become surprising. The participants in the natural group are trained on natural behavior, so their long-term expectations are in line with contextual expectations. Moreover, we hypothesize that intact objects have a very strong long-term prior. Therefore, after training in atypical group, this stubborn prior is modulated to develop weak short-term expectations as shown on the left panel of Figure 3.15. On the other hand, breaking objects have a weak or broad prior. It is possible that with the short-term training in the atypical group, observers develop a strong prior in favor of the atypical behavior because this was a weak prior to start with, as shown in the right panel of Figure 3.15. But how does this trend lead to a reduction in RTs toward the end of the session? We discuss this below:

At the beginning of the training session for the atypical group, when surprising stimuli are presented, the prior and likelihood disagree with each other as the stimulus is inconsistent with long-term priors. Hence, more iterations are required to reach a decision, and participants take longer to respond. Over time, their priors are updated, and they develop short-term or contextual expectations. By the end of the session, their newly developed contextual expectations are consistent with the stimuli presented, and hence they become faster at responding,

leading to shorter RTs. For the Natural group, since they are presented with stimuli that are consistent with their long-term expectations, therefore, we don't see much reduction in their reaction times.

Experimental session: We have explained earlier how our model predicts longer RTs for surprising stimuli. Moreover, our proposed model also has the potential to further explain the interplay between long-term and short-term expectations. Recall that, in our experimental session, we found a significant effect of expectation for intact objects but not for breaking objects in the natural group. Conversely, there was a significant effect for breaking objects but not for intact objects in the atypical group. A broad prior for different types of breaking in natural group may explain why there is little or no effect on reaction times for breaking objects (scenario 3 in Figure 3.16). Whereas, in the natural group, the prior for the intact objects to be rigid may be very strong, i.e. a very sharp prior distribution (scenario 1 in Figure 3.16). In the atypical group, since the observers develop a strong prior in favor of the atypical behavior of breaking objects, therefore we observe a strong effect on the RTs for breaking objects (scenario 2 in Figure 3.16). For the intact object in an atypical group, the situation could be different. Observers developed a weak prior in training, hence the decision takes shorter (scenario 4 in Figure 3.16).

This Bayesian framework is very promising and can provide important insights into the effects of expectations on low-level perceptual decisions regarding dynamic properties of material perception and can shed light on the interaction between long-term and short-term expectations.

3.3.4 Conclusion

To conclude, we found that unmet expectations about dynamic material properties delay perceptual decisions. We argue that high-level expectations about material properties affect relatively low-level perceptual processes even when those expectations are not directly task-relevant. Furthermore, we show that through training, participants form new context-dependent expectations. Those

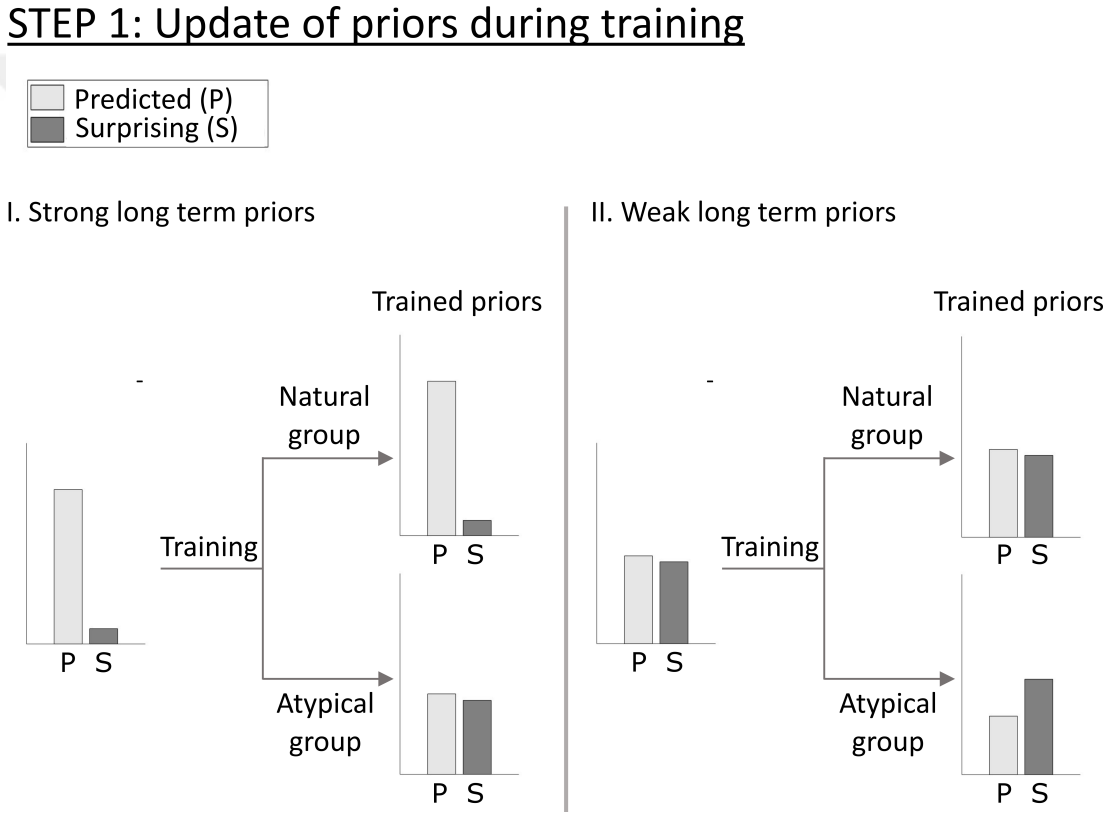

Figure 3.15: This figure shows the first step where long-term priors are updated through training. The left panel shows modulation of strong long-term priors for both groups and the right panel shows modulation of weak long-term priors for both groups. Here predicted stimuli include shattering and rigid objects, while surprising stimuli include graveling and wobbling objects.

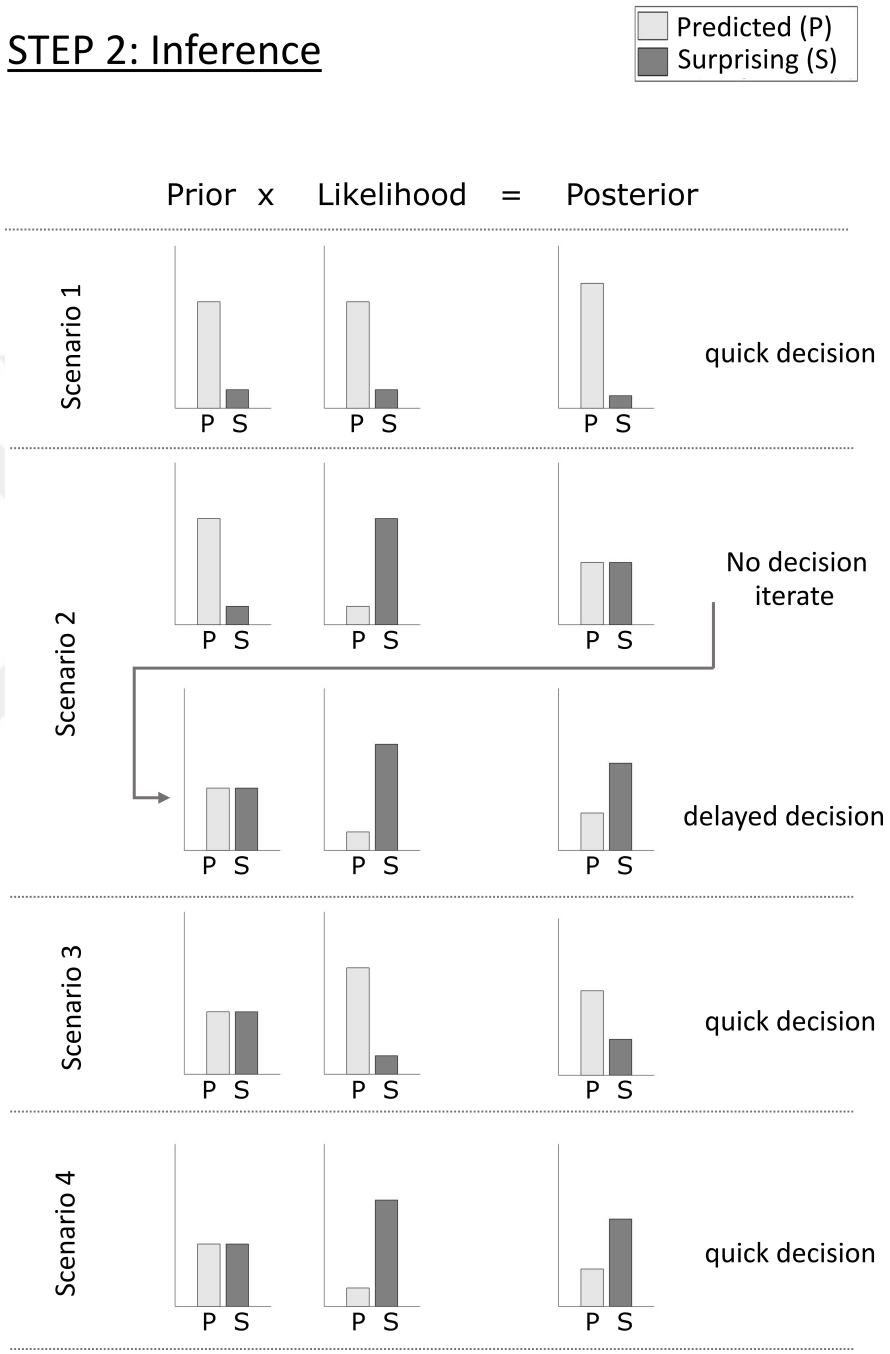

Figure 3.16: This figure shows the second step that how an iterative model may be able to predict the empirical results. 4 possible scenarios based on the strength of priors are depicted (explained in the discussion section). Here, the predicted condition can be shattering and rigid objects for the natural group and can be graveling and wobbling objects for the atypical group. While surprising can be graveling and wobbling for the natural group and can be shattering and rigid for the atypical group.

newly formed context-dependent expectations and long-term expectations together shape the perceptual processes, which can be formulated using a Bayesian updating approach.

3.4 Future Directions

Testing the proposed Bayesian model: The Bayesian model proposed in the current study is a very promising model to explain the interplay between long-term and short-term expectations. However, this model has not been tested on the empirical data. In the future, we plan to test this model against the empirical data.

Neural correlates of expectations. As discussed earlier, the literature on neural activity corresponding to expectations is inconsistent. Electrophysiological studies report an increase in neural activity with the expected stimuli, and fMRI studies report the otherwise i.e., expectation suppression [168, 238, 240, 241]. Moreover, there are fMRI studies investigating expectations in object perception, but neural correlates of expectations in the perception of dynamic material properties have not been studied [281]. Hence, an fMRI study of the current experiment can shed light on the neural correlates of expectations in the perception of dynamic material properties. Investigating how the activity of regions are modulated before and after a training session and comparing natural and atypical group can also provide important insights into how expectations are formed and what is the interplay between long-term and short-term expectations.

Mechanistic cortical model. The current thesis proposes an iterative Bayesian framework to model the effect of expectations on dynamic material properties. Such a computational model provides an abstraction of the underlying mechanisms. Further investigations are needed to understand how such a framework is implemented in the brain. Mechanistic cortical models have the potential to provide a powerful tool for understanding the mechanisms underlying Bayesian processing [248]. Therefore testing the empirical data with such models can be very useful.

Effects of expectations on a broader set of material properties: In the current study, we investigate the role of expectations about dynamic material

properties on low-level perceptual decisions. However, we focused on only a single property, i.e., *fragility*, by asking a very specific question *'Does it break?'*. It will be interesting to know the impact of expectations on low-level perceptual decisions by considering other dynamic material properties. Moreover, investigating what dimensions of material properties and which cues an observer accesses when judging the future outcome will shed light on the extent to which expectations influence the low-level perception.

Interplay between long-term and short-term expectations: Here, we find that the stronger long-term expectations are, the more difficult it is to modulate these priors through training to develop short-term expectations. Therefore, the short-term expectations they develop are weaker if their long-term expectations are strong. But there remain several open questions. Our training session consisted of only a few trials per condition. Will participants develop stronger short-term expectations if the duration of training is increased? Therefore, further research is needed to investigate an interaction between the time/strength of training on the modulation of long-term expectations. Moreover, our experimental session followed immediately after the training session. It will be interesting to investigate how long participants will be able to retain contextual expectations by testing them subsequently on the following days. Finally, we established the interaction between the strength of short-term and long-term expectations, but what are the plausible reasons behind some of the long-term expectations being stronger than others is a question that needs to be addressed.

Virtual reality experiments: The cues we draw for accessing material properties may vary depending on the task. For instance, whether we are predicting the future state of objects or we are planning an interaction with them, or recognizing the material properties of objects. Here we only investigate the effect of expectations on relatively low-level perceptual decisions by asking a simple question, *'Does it break?'*. In the real world, we are continually interacting with objects, So how expectations influence our interactions with objects can be studied through experiments involving virtual reality, where participants plan actions and interact with objects.

Chapter 4

General Discussion

This thesis emphasizes the fundamental role context plays in visual perception to disambiguate noisy and incomplete information and to optimize behavior. However, despite the acknowledged importance of contextual effects, there remain several open questions in the existing literature that require further research. The present study seeks to address this by investigating spatial and cognitive contextual effects in visual processing in two separate lines of work. I discussed the specific implications of the results of each of the two studies in detail under the discussion sections of the respective chapters. Here, I discuss the broader implications of the research.

The study presented in Chapter 2 highlights the importance of spatial context in visual perception. Spatial context is especially an interesting topic for visual perception because the visual areas are topographically arranged. Neurons present at different topographical locations interact with each other to modulate their responses. Therefore, investigating the effects of spatial context on the processing of visual information is very helpful in understanding these interactions and functional relationships between different topographical locations in the early visual cortex. Here, we study a low-level visual feature (lightness). Low-level visual features lay the foundation of overall visual perception. Hence, it is essential to understand how the visual system extracts this low-level information from the

visual field. Our study provides evidence that neural correlates of contextual processing of lightness are reflected in the early visual cortex. In line with our study, several studies have reported the spatial contextual effects on low-level visual features to be correlated with neural activity in early visual areas. For instance, the spatial context in orientation perception gives rise to the tilt illusion as shown in Figure 2.1-C. The investigation of neural correlates of this illusion demonstrates that the activity of early visual cortices correlates with tilt illusion [282]. Similarly, research on size constancy demonstrating the effect of spatial context reports early visual areas involved in such illusions [283, 284]. Contextual effects on contrast perception are also reported to correlate with the activity of the primary visual cortex [285]. Overall, this reflects that spatial context influences most low-level visual features in the early stages of visual processing. The involvement of V1 in these contextual effects shows that the processing of information in V1 is much more complex, where interaction between features present inside the receptive field with those present outside at a considerable distance from the receptive field determines the responses. However, the involvement of early visual areas in the contextual processing of low-level visual features does not rule out the possibility of top-down influences from higher-level areas, which is discussed in the following paragraphs.

In Chapter 3, we present our study on the effects of expectations on the perception of dynamic material properties and low-level perceptual decisions, highlighting the importance of cognitive context. Understanding the role of expectations within a cognitive context provides insights into the complex interplay between cognition, perception, and behavior. Our study shows that low-level perceptual decisions are delayed when expectations about dynamic material properties are unmet. We also show that cognitive context is not static. It is rather dynamic by demonstrating that long-term expectations about dynamic material properties can be altered by training to form short-term contextual expectations. The aim of the current study was not focused on finding the neural correlates of cognitive contextual effects, but such effects are shown to activate the memory representation of the relevant objects and their learned associations, including what properties they hold in a given dynamic context [176]. Hence, such cognitive influences come

into play at a higher level of visual processing.

Thus contextual effects can manifest themselves either at a low level or a high level in the visual hierarchy, depending on the kind of context. For instance, the contextual processing of lightness is low level. It relies more on the bottom-up processing of information, where the spatial context derived from the incoming visual information affects the perception. In contrast, the effect of cognitive context in the form of prior knowledge comes into play at a higher level of visual processing and relies on top-down processing.

It is also interesting to note that the effect of spatial context in lightness perception study is unconscious. Interestingly, knowledge about the visual illusions that are perceived because of the spatial context does not make them disappear, even if you design them yourself. For instance, I know that Figure 2.1-E has a disk of constant luminance embedded in a different surround, but I still perceive it as having a different luminance. The same is true for all the illusions in Figure 2.1 demonstrating the effect of spatial context. On the other hand, the effect of cognitive context in the current study is conscious. When participants are presented with stimuli that violate their expectations, they are aware of it. This conscious experience also leads them to develop contextual or short-term expectations. This is intriguing that different types of contextual effects can be operated at different levels of conscious awareness at different levels in the visual hierarchy.

This discussion, however, does not imply that bottom-up and top-down effects are entirely distinct. Some studies imply an interaction between bottom-up processes with top-down effects in the processing of low-level visual features. For instance, several studies support that high-level areas modulate lightness through top-down effects by showing that the activity of higher-level areas correlates with perceived lightness [116, 77]. Similarly, attention is reported to modulate many of the spatial contextual effects through top-down processes, e.g., in size perception and contrast perception [286, 287]. However, the involvement of attention specifically in lightness perception is contradictory [288, 289, 290].

The role of expectations in lightness perception has also been a topic of debate recently. A prominent study by Levin et al. [291] demonstrated that expectations about skin tone could affect the perceived lightness, showing that observers perceive white faces to be whiter than black faces. Firestone et al. [292] repeated the experiment and argued for the possibility of low-level features instead of expectations giving rise to differences in lightness perception. They proposed it to be merely a brightness effect. However, many studies later demonstrated the effect of face race on the perceived lightness of the face, now known as the 'face race lightness illusion' [293, 294]. Nonetheless, this illusion provides compelling evidence for the role of cognitive context in processing low-level visual features. To what extent cognitive context can penetrate low-level vision is still a debate and needs to be addressed further [295].

In conclusion, context can have a range of influences on visual perception, from affecting the appearance of a simple visual stimulus to the perception of learned associations in object perception and perception of material properties. Investigating these contextual effects is really essential to understand how visual information is processed in general, which in turn can help us to develop accurate models of visual processing. Not only this, but it can also shed light on the broader cognitive processes such as attention and expectations, and decision-making, which can have important implications in artificial intelligence, computer vision, and image processing. Contextual processing is also affected in many neurological disorders, including Autism spectrum disorder (ASD), Schizophrenia, Attention deficit hyperactivity disorder (ADHD), and Alzheimer's disease. The knowledge of neural circuits involved in contextual processing can help develop interventions for these neurological disorders. Moreover, understanding the effects of expectations and contextual processing, especially in object and material perception, can also be useful for designers and advertisement agencies, where they can create effective marketing strategies using this knowledge. Hence, the current thesis plays an essential role in the literature by broadening our understanding of contextual effects on visual perception and providing a direction for future research in the area.

Bibliography

[1] W. J. Ma, J. M. Beck, P. E. Latham, and A. Pouget, "Bayesian inference with probabilistic population codes," *Nature neuroscience*, vol. 9, no. 11, pp. 1432–1438, 2006.

[2] S. Maren, K. L. Phan, and I. Liberzon, "The contextual brain: implications for fear conditioning, extinction and psychopathology," *Nature reviews neuroscience*, vol. 14, no. 6, pp. 417–428, 2013.

[3] H. Ebbinghaus, "Grundzüge der psychologie volumes i and ii," *Leipzig: Verlag von Viet & Co*, 1902.

[4] R. Over, J. Broerse, and B. Crassini, "Orientation illusion and masking in central and peripheral vision.," *Journal of experimental psychology*, vol. 96, no. 1, p. 25, 1972.

[5] F. Kingdom, "Simultaneous contrast: the legacies of hering and helmholtz," 1997.

[6] H. K. Walker, W. D. Hall, and J. W. Hurst, *Clinical methods: the history, physical, and laboratory examinations*. Butterworths, 1990.

[7] T. Kunnapas, "Distance perception as a function of available visual cues.," *Journal of experimental psychology*, vol. 77, no. 4, p. 523, 1968.

[8] D. Todorović, "Context effects in visual perception and their explanations," *Review of psychology*, vol. 17, no. 1, pp. 17–32, 2010.

[9] J. S. Girgus, S. Coren, and M. Agdern, "The interrelationship between the ebbinghaus and delboeuf illusions.," *Journal of Experimental Psychology*, vol. 95, no. 2, p. 453, 1972.

[10] G. Y. Yildiz, I. Sperandio, C. Kettle, and P. A. Chouinard, "A review on various explanations of ponzo-like illusions," *Psychonomic Bulletin & Review*, pp. 1–28, 2022.

[11] J. J. Gibson, "Adaptation, after-effect, and contrast in the perception of tilted lines. ii. simultaneous contrast and the areal restriction of the after-effect.," *Journal of Experimental Psychology*, vol. 20, no. 6, p. 553, 1937.

[12] P. Wenderoth and S. Johnstone, "The different mechanisms of the direct and indirect tilt illusions," *Vision research*, vol. 28, no. 2, pp. 301–312, 1988.

[13] G. Westheimer, "Illusions in the spatial sense of the eye: Geometrical–optical illusions and the neural representation of space," *Vision research*, vol. 48, no. 20, pp. 2128–2142, 2008.

[14] R. J. Snowden and S. T. Hammett, "The effects of surround contrast on contrast thresholds, perceived contrast and contrast discrimination," *Vision research*, vol. 38, no. 13, pp. 1935–1945, 1998.

[15] J. Xing and D. J. Heeger, "Center-surround interactions in foveal and peripheral vision," *Vision research*, vol. 40, no. 22, pp. 3065–3072, 2000.

[16] C. Chubb, G. Sperling, and J. A. Solomon, "Texture interactions determine perceived contrast.," *Proceedings of the National Academy of Sciences*, vol. 86, no. 23, pp. 9631–9635, 1989.

[17] M. W. Cannon and S. C. Fullenkamp, "Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations," *Vision research*, vol. 31, no. 11, pp. 1985–1998, 1991.

[18] A. Gilchrist, *Seeing black and white*. Oxford University Press, 2006.

[19] A. L. Gilchrist, "Lightness and brightness," *Current Biology*, vol. 17, no. 8, pp. R267–R269, 2007.

[20] E. H. Adelson, "24 lightness perception and lightness illusions," 2000.

[21] M. White, "A new effect of pattern on perceived lightness," *Perception*, vol. 8, no. 4, pp. 413–416, 1979.

[22] W. Benary, "Beobachtungen zu einem experiment über helligkeitskontrast," *Psychologische Forschung*, vol. 5, no. 1, pp. 131–142, 1924.

[23] A. F. Rossi and M. A. Paradiso, "Temporal limits of brightness induction and mechanisms of brightness perception," *Vision research*, vol. 36, no. 10, pp. 1391–1398, 1996.

[24] A. F. Rossi and M. A. Paradiso, "Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex," *Journal of Neuroscience*, vol. 19, no. 14, pp. 6145–6156, 1999.

[25] H. Komatsu, "The neural mechanisms of perceptual filling-in," *Nature reviews neuroscience*, vol. 7, no. 3, pp. 220–231, 2006.

[26] J.-D. Haynes, R. B. Lotto, and G. Rees, "Responses of human visual cortex to uniform surfaces," *Proceedings of the National Academy of Sciences*, vol. 101, no. 12, pp. 4286–4291, 2004.

[27] B. Blakeslee and M. E. McCourt, "Similar mechanisms underlie simultaneous brightness contrast and grating induction," *Vision research*, vol. 37, no. 20, pp. 2849–2869, 1997.

[28] P. Sinha, S. Crucilla, T. Gandhi, D. Rose, A. Singh, S. Ganesh, U. Mathur, and P. Bex, "Mechanisms underlying simultaneous brightness contrast: early and innate," *Vision research*, vol. 173, pp. 41–49, 2020.

[29] M. Pereverzeva and S. O. Murray, "Neural activity in human v1 correlates with dynamic lightness induction," *Journal of Vision*, vol. 8, no. 15, pp. 8–8, 2008.

[30] F. W. Cornelissen, A. R. Wade, T. Vladusich, R. F. Dougherty, and B. A. Wandell, "No functional magnetic resonance imaging evidence for brightness and color filling-in in early human visual cortex," *Journal of Neuroscience*, vol. 26, no. 14, pp. 3634–3641, 2006.

[31] E. H. Adelson, "Perceptual organization and the judgment of brightness," *Science*, vol. 262, no. 5142, pp. 2042–2044, 1993.

[32] J. P. Richter *et al.*, *The notebooks of Leonardo da Vinci*, vol. 2. Courier Corporation, 1970.

[33] H. von Helmholtz, "Handbook of physiological optics," *Handbuch der Physiologischen Optik (German). Hamburg: Voss*, 1866.

[34] A. Gelb, *Die „Farbenkonstanz "der Sehdinge*. Springer, 1929.

[35] E. Hering, *Outlines of a Theory of the Light Sense*. Harvard University Press, 1964.

[36] A. L. Gilchrist, "Perceived lightness depends on perceived spatial arrangement," *Science*, vol. 195, no. 4274, pp. 185–187, 1977.

[37] H. K. Hartline, "The receptive fields of optic nerve fibers," *American Journal of Physiology-Legacy Content*, vol. 130, no. 4, pp. 690–699, 1940.

[38] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, vol. 160, no. 1, p. 106, 1962.

[39] C. Enroth-Cugell and J. G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat," *The Journal of physiology*, vol. 187, no. 3, pp. 517–552, 1966.

[40] S. E. Palmer, *Vision science: Photons to phenomenology*. MIT press, 1999.

[41] E. W. Yund, D. Snodderly, N. Hepler, and R. DeValois, "Brightness contrast effects in monkey lateral geniculate nucleus," *Sensory Processes*, vol. 1, no. 3, pp. 260–271, 1977.

[42] R. L. DeValois, P. Burke, K. K. De Valois, and K. K. DeValois, *Spatial vision*. Oxford University Press on Demand, 1990.

[43] R. L. De Valois and P. L. Pease, "Contours and contrast: responses of monkey lateral geniculate nucleus cells to luminance and color figures," *Science*, vol. 171, no. 3972, pp. 694–696, 1971.

[44] E. W. Yund and J. C. Armington, "Color and brightness contrast effects as a function of spatial variables," *Vision Research*, vol. 15, no. 8-9, pp. 917–929, 1975.

[45] D. Jameson, "Opponent-colours theory in the light of physiological findings," in *Central and Peripheral Mechanisms of Colour Vision: Proceedings of an International Symposium Held at The Wenner-Gren Center Stockholm, June 14–15, 1984*, pp. 83–102, Springer, 1985.

[46] D. Jameson and L. M. Hurvich, "Essay concerning color constancy," *Annual review of psychology*, vol. 40, no. 1, pp. 1–24, 1989.

[47] F. Kingdom and B. Moulden, "A multi-channel approach to brightness coding," *Vision Research*, vol. 32, no. 8, pp. 1565–1582, 1992.

[48] B. Blakeslee and M. E. McCourt, "A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction," *Vision research*, vol. 39, no. 26, pp. 4361–4377, 1999.

[49] B. Blakeslee and M. E. McCourt, "A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization," *Vision research*, vol. 44, no. 21, pp. 2483–2503, 2004.

[50] B. Blakeslee, W. Pasieka, and M. E. McCourt, "Oriented multiscale spatial filtering and contrast normalization: a parsimonious model of brightness induction in a continuum of stimuli including white, howe and simultaneous brightness contrast," *Vision Research*, vol. 45, no. 5, pp. 607–615, 2005.

[51] D. Corney and R. B. Lotto, "What are lightness illusions and why do we see them?," *PLoS computational biology*, vol. 3, no. 9, p. e180, 2007.

[52] G. L. Walls, "The filling-in process," *Optometry and Vision Science*, vol. 31, no. 7, pp. 329–341, 1954.

[53] H. Gerrits and A. Vendrik, "Simultaneous contrast, filling-in process and information processing in man's visual system," *Experimental Brain Research*, vol. 11, pp. 411–430, 1970.

[54] H. Gerrits, B. De Haan, and A. Vendrik, "Experiments with retinal stabilized images. relations between the observations and neural data," *Vision research*, vol. 6, no. 7-8, pp. 427–440, 1966.

[55] J. Larimer and T. Piantanida, "The impact of boundaries on color: Stabilized image studies," in *Image Processing, Analysis, Measurement, and Quality*, vol. 901, pp. 241–247, SPIE, 1988.

[56] J. Krauskopf, "Effect of retinal image stabilization on the appearance of heterochromatic targets," *JOSA*, vol. 53, no. 6, pp. 741–744, 1963.

[57] A. L. Yarbus and A. L. Yarbus, "Eye movements during perception of complex objects," *Eye movements and vision*, pp. 171–211, 1967.

[58] M. A. Paradiso and K. Nakayama, "Brightness perception and filling-in," *Vision research*, vol. 31, no. 7-8, pp. 1221–1236, 1991.

[59] M. Davey, T. Maddess, and M. Srinivasan, "The spatiotemporal properties of the craik–o'brien–cornsweet effect are consistent with 'filling-in'," *Vision Research*, vol. 38, no. 13, pp. 2037–2046, 1998.

[60] K. Hamburger, H. Prior, V. Sarris, and L. Spillmann, "Filling-in with colour: Different modes of surface completion," *Vision research*, vol. 46, no. 6-7, pp. 1129–1138, 2006.

[61] M. A. Paradiso and S. Hahn, "Filling-in percepts produced by luminance modulation," *Vision research*, vol. 36, no. 17, pp. 2657–2663, 1996.

[62] B. Blakeslee and M. E. McCourt, "Nearly instantaneous brightness induction," *Journal of vision*, vol. 8, no. 2, pp. 15–15, 2008.

[63] G. Barkeley, "An essay towards a new theory of vision ebook," *In Works*, vol. 1, 1709.

[64] G. Barkeley, "Three dialogues between hylas and philonous," *In Works*, vol. 2, 1710.

[65] D. Katz, *The world of colour*. Routledge, 2013.

[66] H. Wallach, "Brightness constancy and the nature of achromatic colors.," *Journal of experimental psychology*, vol. 38, no. 3, p. 310, 1948.

[67] K. Koffka, "Principles of gestalt psychology. publ," *Lund Humphries, London*, 1935.

[68] W. Köhler, *Gestalt psychology: An introduction to new concepts in modern psychology*, vol. 18. WW Norton & Company, 1970.

[69] G. Katona, "Zur analyse der helligkeitskonstanz.," *Psychologische Forschung*, 1929.

[70] W. Wolff, "Über die kontrasterregende wirkung der transformierten farben," *Psychologische Forschung*, vol. 18, pp. 90–97, 1933.

[71] L. Kardos, "Ding und schatten. eine experimentelle untersuchung über die grundlagen des farbensehens.," *Zeitschrift für Psychologie und Physiologie der Sinnesorgane. Abt. 1. Zeitschrift für Psychologie*, 1934.

[72] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Josa*, vol. 61, no. 1, pp. 1–11, 1971.

[73] L. Arend, J. Buehler, and G. Lockhead, "Difference information in brightness perception," *Perception & Psychophysics*, vol. 9, no. 3, pp. 367–370, 1971.

[74] P. Whittle and P. Challands, "The effect of background luminance on the brightness of flashes," *Vision research*, vol. 9, no. 9, pp. 1095–1110, 1969.

[75] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou, "An anchoring theory of lightness perception.," *Psychological review*, vol. 106, no. 4, p. 795, 1999.

[76] R. F. Murray, "Lightness perception in complex scenes," *Annual Review of Vision Science*, vol. 7, pp. 417–436, 2021.

[77] A. Perna, M. Tosetti, D. Montanaro, and M. C. Morrone, "Neuronal mechanisms for illusory brightness perception in humans," *Neuron*, vol. 47, no. 5, pp. 645–651, 2005.

[78] R. Von Der Heydt, H. S. Friedman, and H. Zhou, "Searching for the neural mechanisms of color filling-in," *Filling-in: From perceptual completion to cortical reorganization*, pp. 106–127, 2003.

[79] M. Pereverzeva and W. D. Bromfield, "Spatial nonuniformities and velocity of filling-in in dynamic brightness induction," *Journal of Vision*, vol. 13, no. 5, pp. 17–17, 2013.

[80] A. E. Robinson and V. R. de Sa, "Brief presentations reveal the temporal dynamics of brightness induction and white's illusion," *Vision research*, vol. 48, no. 22, pp. 2370–2381, 2008.

[81] T. N. Cornsweet, "Visual perception academic press," *New York London*, 1970.

[82] M. Davidson, "Perturbation approach to spatial brightness interaction in human vision," *JOSA*, vol. 58, no. 9, pp. 1300–1308, 1968.

[83] B. Blakeslee and M. E. McCourt, "Brightness induction magnitude declines with increasing distance from the inducing field edge," *Vision research*, vol. 78, pp. 39–45, 2013.

[84] C. W. Tyler and J. Torres, "Frequency response characteristics for sinusoidal movement in the fovea and periphery," *Perception & Psychophysics*, vol. 12, pp. 232–236, 1972.

[85] S. P. McKee and D. G. Taylor, "Discrimination of time: comparison of foveal and peripheral sensitivity," *JOSA A*, vol. 1, no. 6, pp. 620–627, 1984.

[86] R. L. De Valois, M. A. Webster, K. K. De Valois, and B. Lingelbach, "Temporal properties of brightness and color induction," *Vision research*, vol. 26, no. 6, pp. 887–897, 1986.

[87] Q. Zaidi, B. Yoshimi, N. Flanigan, and A. Canova, "Lateral interactions within color mechanism in simultaneous induced contrast," *Vision research*, vol. 32, no. 9, pp. 1695–1707, 1992.

[88] A. F. Rossi, C. D. Rittenhouse, and M. A. Paradiso, "The representation of brightness in primary visual cortex," *Science*, vol. 273, no. 5278, pp. 1104–1107, 1996.

[89] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of physiology*, vol. 148, no. 3, p. 574, 1959.

[90] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of physiology*, vol. 195, no. 1, pp. 215–243, 1968.

[91] V. A. Lamme, "The neurophysiology of figure-ground segregation in primary visual cortex," *Journal of neuroscience*, vol. 15, no. 2, pp. 1605–1615, 1995.

[92] W. Li, V. Piëch, and C. D. Gilbert, "Contour saliency in primary visual cortex," *Neuron*, vol. 50, no. 6, pp. 951–962, 2006.

[93] A. M. Slllito, K. L. Grieve, H. E. Jones, J. Cudeiro, and J. Davls, "Visual cortical mechanisms detecting focal orientation discontinuities," *Nature*, vol. 378, no. 6556, pp. 492–496, 1995.

[94] D. H. Baker and E. W. Graf, "Contextual effects in speed perception may occur at an early stage of processing," *Vision research*, vol. 50, no. 2, pp. 193–201, 2010.

[95] S. Kastner, H.-C. Nothdurft, and I. N. Pigarev, "Neuronal correlates of pop-out in cat striate cortex," *Vision research*, vol. 37, no. 4, pp. 371–376, 1997.

[96] V. A. Lamme, H. Super, and H. Spekreijse, "Feedforward, horizontal, and feedback processing in the visual cortex," *Current opinion in neurobiology*, vol. 8, no. 4, pp. 529–535, 1998.

[97] J. B. Levitt and J. S. Lund, "Contrast dependence of contextual effects in primate visual cortex," *Nature*, vol. 387, no. 6628, pp. 73–76, 1997.

[98] F. Sengpiel, A. Sen, and C. Blakemore, "Characteristics of surround inhibition in cat area 17," *Experimental brain research*, vol. 116, pp. 216–228, 1997.

[99] D. C. Somers, E. V. Todorov, A. G. Siapas, L. J. Toth, D.-S. Kim, and M. Sur, "A local circuit approach to understanding integration of long-range inputs in primary visual cortex.," *Cerebral cortex (New York, NY: 1991)*, vol. 8, no. 3, pp. 204–217, 1998.

[100] M. K. Kapadia, G. Westheimer, and C. D. Gilbert, "Dynamics of spatial summation in primary visual cortex of alert monkeys," *Proceedings of the National Academy of Sciences*, vol. 96, no. 21, pp. 12073–12078, 1999.

[101] M. P. Sceniak, D. L. Ringach, M. J. Hawken, and R. Shapley, "Contrast's effect on spatial summation by macaque v1 neurons," *Nature neuroscience*, vol. 2, no. 8, pp. 733–739, 1999.

[102] J. R. Cavanaugh, W. Bair, and J. A. Movshon, "Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons," *Journal of neurophysiology*, vol. 88, no. 5, pp. 2530–2546, 2002.

[103] O. Sadakane, H. Ozeki, T. Naito, T. Akasaki, T. Kasamatsu, and H. Sato, "Contrast-dependent, contextual response modulation in primary visual cortex and lateral geniculate nucleus of the cat," *European Journal of Neuroscience*, vol. 23, no. 6, pp. 1633–1642, 2006.

[104] J. M. Ichida, L. Schwabe, P. C. Bressloff, and A. Angelucci, "Response facilitation from the "suppressive" receptive field surround of macaque v1 neurons," *Journal of neurophysiology*, vol. 98, no. 4, pp. 2168–2181, 2007.

[105] L. Schwabe, J. M. Ichida, S. Shushruth, P. Mangapathy, and A. Angelucci, "Contrast-dependence of surround suppression in macaque v1: experimental testing of a recurrent network model," *Neuroimage*, vol. 52, no. 3, pp. 777–792, 2010.

[106] M. E. McCourt, "A spatial frequency dependent grating-induction effect," *Vision Research*, vol. 22, no. 1, pp. 119–134, 1982.

[107] W. M. Maguire and J. S. Baizer, "Luminance coding of briefly presented stimuli in area 17 of the rhesus monkey.," *Journal of Neurophysiology*, vol. 47, no. 1, pp. 128–137, 1982.

[108] Y. Kayama, R. R. Riso, J. R. Bartlett, and R. W. Doty, "Luxotonic responses of units in macaque striate cortex," *Journal of Neurophysiology*, vol. 42, no. 6, pp. 1495–1517, 1979.

[109] J. R. Bartlett and R. W. Doty Sr, "Response of units in striate cortex of squirrel monkeys to visual and electrical stimuli.," *Journal of Neurophysiology*, vol. 37, no. 4, pp. 621–641, 1974.

[110] M. Kinoshita and H. Komatsu, "Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex," *Journal of neurophysiology*, vol. 86, no. 5, pp. 2559–2570, 2001.

[111] C. C. Boucard, J. J. van Es, R. P. Maguire, and F. W. Cornelissen, "Functional magnetic resonance imaging of brightness induction in the human visual cortex," *Neuroreport*, vol. 16, no. 12, pp. 1335–1338, 2005.

[112] Y. Sasaki and T. Watanabe, "The primary visual cortex fills in color," *Proceedings of the National Academy of Sciences*, vol. 101, no. 52, pp. 18251–18256, 2004.

[113] H. Boyaci, F. Fang, S. O. Murray, and D. Kersten, "Responses to lightness variations in early human visual cortex," *Current Biology*, vol. 17, no. 11, pp. 989–993, 2007.

[114] C. P. Hung, B. M. Ramsden, L. M. Chen, and A. W. Roe, "Building surfaces from borders in areas 17 and 18 of the cat," *Vision research*, vol. 41, no. 10-11, pp. 1389–1407, 2001.

[115] A. W. Roe, H. D. Lu, and C. P. Hung, "Cortical processing of a brightness illusion," *Proceedings of the National Academy of Sciences*, vol. 102, no. 10, pp. 3869–3874, 2005.

[116] H. Zhou, M. Davidson, P. Kok, L. Y. McCurdy, F. P. de Lange, H. Lau, and K. Sandberg, "Spatiotemporal dynamics of brightness coding in human

visual cortex revealed by the temporal context effect," *Neuroimage*, vol. 205, p. 116277, 2020.

[117] D. M. Eagleman, J. E. Jacobson, and T. J. Sejnowski, "Perceived luminance depends on temporal context," *Nature*, vol. 428, no. 6985, pp. 854–856, 2004.

[118] D. A. Ruff, D. H. Brainard, and M. R. Cohen, "Neuronal population mechanisms of lightness perception," *Journal of Neurophysiology*, vol. 120, no. 5, pp. 2296–2310, 2018.

[119] B. N. Bushnell, P. J. Harding, Y. Kosai, W. Bair, and A. Pasupathy, "Equiluminance cells in visual cortical area v4," *Journal of Neuroscience*, vol. 31, no. 35, pp. 12398–12412, 2011.

[120] A. Malik and H. Boyaci, "Neural correlates of spatiotemporal properties of lightness induction," *bioRxiv*, pp. 2023–07, 2023.

[121] E. G. Heinemann, "Simultaneous brightness induction as a function of inducing-and test-field luminances.," *Journal of experimental psychology*, vol. 50, no. 2, p. 89, 1955.

[122] H. Von Helmholtz, *Handbuch der physiologischen Optik*, vol. 9. Voss, 1867.

[123] I. Alhazen, "Book of optics (a. sabra, trans.)," *The otics of Ibn al-Haytham*, vol. 2, 1883.

[124] E. J. Anderson, S. C. Dakin, and G. Rees, "Monocular signals in human lateral geniculate nucleus reflect the craik–cornsweet–o'brien effect," *Journal of Vision*, vol. 9, no. 12, pp. 14–14, 2009.

[125] S. C. Dakin and P. J. Bex, "Natural image statistics mediate brightness 'filling in'," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 270, no. 1531, pp. 2341–2348, 2003.

[126] S. E. Bouvier, K. S. Cardinal, and S. A. Engel, "Activity in visual area v4 correlates with surface perception," *Journal of vision*, vol. 8, no. 7, pp. 28–28, 2008.

[127] E. Economou, S. Zdravkovic, and A. Gilchrist, "Anchoring versus spatial filtering accounts of simultaneous lightness contrast," *Journal of vision*, vol. 7, no. 12, pp. 2–2, 2007.

[128] H. Boyaci, F. Fang, S. O. Murray, and D. Kersten, "Perceptual grouping-dependent lightness processing in human early visual cortex," *Journal of vision*, vol. 10, no. 9, pp. 4–4, 2010.

[129] L. N. Vinke and S. Ling, "Luminance potentiates human visuocortical responses," *Journal of Neurophysiology*, vol. 123, no. 2, pp. 473–483, 2020.

[130] W. S. Geisler, D. G. Albrecht, and A. M. Crane, "Responses of neurons in primary visual cortex to transient changes in local contrast and luminance," *Journal of Neuroscience*, vol. 27, no. 19, pp. 5063–5067, 2007.

[131] D. H. Brainard and S. Vision, "The psychophysics toolbox," *Spatial vision*, vol. 10, no. 4, pp. 433–436, 1997.

[132] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[133] RStudio Team, *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA, 2020.

[134] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[135] R. V. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2023. R package version 1.8.4-1.

[136] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[137] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[138] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[139] A. Remole, "Spatial characteristics of the border enhancement region in a flickering field," *Optometry and Vision Science*, vol. 47, no. 10, pp. 779–786, 1970.

[140] S. P. MacEvoy and M. A. Paradiso, "Lightness constancy in primary visual cortex," *Proceedings of the National Academy of Sciences*, vol. 98, no. 15, pp. 8827–8831, 2001.

[141] A. F. Rossi, R. Desimone, and L. G. Ungerleider, "Contextual modulation in primary visual cortex of macaques," *Journal of Neuroscience*, vol. 21, no. 5, pp. 1698–1709, 2001.

[142] B. G. Goodyear and R. S. Menon, "Effect of luminance contrast on bold fmri response in human primary visual areas," *Journal of Neurophysiology*, vol. 79, no. 4, pp. 2204–2207, 1998.

[143] Y. Chai, D. A. Handwerker, S. Marrett, J. Gonzalez-Castillo, E. P. Merriam, A. Hall, P. J. Molfese, and P. A. Bandettini, "Visual temporal frequency preference shows a distinct cortical architecture using fmri," *Neuroimage*, vol. 197, pp. 13–23, 2019.

[144] C. P. Hung, B. M. Ramsden, and A. W. Roe, "A functional circuitry for edge-induced brightness perception," *Nature neuroscience*, vol. 10, no. 9, pp. 1185–1190, 2007.

[145] V. van de Ven, B. Jans, R. Goebel, and P. De Weerd, "Early human visual cortex encodes surface brightness induced by dynamic context," *Journal of Cognitive Neuroscience*, vol. 24, no. 2, pp. 367–377, 2012.

[146] V. R. Salmela and S. Vanni, "Brightness and transparency in the early visual cortex," *Journal of vision*, vol. 13, no. 7, pp. 16–16, 2013.

[147] A. T. Smith, K. D. Singh, A. L. Williams, and M. W. Greenlee, "Estimating receptive field size from fmri data in human striate and extrastriate visual cortex," *Cerebral cortex*, vol. 11, no. 12, pp. 1182–1190, 2001.

[148] J. A. van Dijk, B. de Haas, C. Moutsiana, and D. S. Schwarzkopf, "Intersession reliability of population receptive field estimates," *NeuroImage*, vol. 143, pp. 293–303, 2016.

[149] K. T. Mullen, B. Thompson, and R. F. Hess, "Responses of the human visual cortex and lgn to achromatic and chromatic temporal modulations: an fmri study," *Journal of vision*, vol. 10, no. 13, pp. 13–13, 2010.

[150] D. V. D'Souza, T. Auer, H. Strasburger, J. Frahm, and B. B. Lee, "Temporal frequency and chromatic processing in humans: an fmri study of the cortical visual areas," *Journal of Vision*, vol. 11, no. 8, pp. 8–8, 2011.

[151] M. E. McCourt and J. J. Foxe, "Brightening prospects for early cortical coding of perceived luminance: a high-density electrical mapping study," *Neuroreport*, vol. 15, no. 1, pp. 49–56, 2004.

[152] D. Marr and W. Vision, "Freeman," *San Francisco*, pp. 41–98, 1982.

[153] D. C. Van Essen and J. H. Maunsell, "Hierarchical organization and functional streams in the visual cortex," *Trends in neurosciences*, vol. 6, pp. 370–375, 1983.

[154] D. Mumford, "Neuronal architectures for pattern-theoretic problems," *Large-scale neuronal theories of the brain*, pp. 125–152, 1994.

[155] E. Sharon, M. Galun, D. Sharon, R. Basri, and A. Brandt, "Hierarchy and adaptivity in segmenting visual scenes," *Nature*, vol. 442, no. 7104, pp. 810–813, 2006.

[156] K. Friston, "A theory of cortical responses," *Philosophical transactions of the Royal Society B: Biological sciences*, vol. 360, no. 1456, pp. 815–836, 2005.

[157] A. Angelucci and J. Bullier, "Reaching beyond the classical receptive field of v1 neurons: horizontal or feedback axons?," *Journal of Physiology-Paris*, vol. 97, no. 2-3, pp. 141–154, 2003.

[158] P.-A. Salin and J. Bullier, "Corticocortical connections in the visual system: structure and function," *Physiological reviews*, vol. 75, no. 1, pp. 107–154, 1995.

[159] K. Doya, S. Ishii, A. Pouget, and R. P. Rao, *Bayesian brain: Probabilistic approaches to neural coding*. MIT press, 2007.

[160] H. Feldman and K. J. Friston, "Attention, uncertainty, and free-energy," *Frontiers in human neuroscience*, vol. 4, p. 215, 2010.

[161] D. Kersten, P. Mamassian, and A. Yuille, "Object perception as bayesian inference," *Annu. Rev. Psychol.*, vol. 55, pp. 271–304, 2004.

[162] T. S. Lee and D. Mumford, "Hierarchical bayesian inference in the visual cortex," *JOSA A*, vol. 20, no. 7, pp. 1434–1448, 2003.

[163] J. H. Maunsell and E. P. Cook, "The role of attention in visual processing," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 357, no. 1424, pp. 1063–1072, 2002.

[164] C. D. Gilbert and T. N. Wiesel, "The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat," *Vision research*, vol. 30, no. 11, pp. 1689–1701, 1990.

[165] P. Vuilleumier and J. Driver, "Modulation of visual processing by attention and emotion: windows on causal interactions between human brain regions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1481, pp. 837–855, 2007.

[166] J. H. Reynolds and L. Chelazzi, "Attentional modulation of visual processing," *Annu. Rev. Neurosci.*, vol. 27, pp. 611–647, 2004.

[167] R. Desimone, "Neural mechanisms for visual memory and their role in attention," *Proceedings of the National Academy of Sciences*, vol. 93, no. 24, pp. 13494–13499, 1996.

[168] C. Summerfield and E. Koechlin, "A neural representation of prior information during perceptual inference," *Neuron*, vol. 59, pp. 336–47, 08 2008.

[169] I. Biederman, "Perceiving real-world scenes," *Science*, vol. 177, no. 4043, pp. 77–80, 1972.

[170] t. E. Palmer, "The effects of contextual scenes on the identification of objects," *Memory & cognition*, vol. 3, no. 5, pp. 519–526, 1975.

[171] J. L. Davenport and M. C. Potter, "Scene consistency in object and background perception," *Psychological science*, vol. 15, no. 8, pp. 559–564, 2004.

[172] A. Torralba, "Modeling global scene factors in attention," *JOSA A*, vol. 20, no. 7, pp. 1407–1418, 2003.

[173] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search.," *Psychological review*, vol. 113, no. 4, p. 766, 2006.

[174] D. Hoiem, A. A. Efros, and M. Hebert, "Putting objects in perspective," *International Journal of Computer Vision*, vol. 80, pp. 3–15, 2008.

[175] J. M. Henderson, P. A. Weeks Jr, and A. Hollingworth, "The effects of semantic consistency on eye movements during complex scene viewing.," *Journal of experimental psychology: Human perception and performance*, vol. 25, no. 1, p. 210, 1999.

[176] A. Friedman, "Framing pictures: the role of knowledge in automatized encoding and memory for gist.," *Journal of experimental psychology: General*, vol. 108, no. 3, p. 316, 1979.

[177] M. E. Auckland, K. R. Cave, and N. Donnelly, "Nontarget objects can influence perceptual processes during object recognition," *Psychonomic Bulletin and Review*, vol. 14, no. 2, p. 332, 2007.

[178] A. Hollingworth, "Does consistent scene context facilitate object perception?," *Journal of experimental psychology: General*, vol. 127, no. 4, p. 398, 1998.

[179] R. M. Warren, "Perceptual restoration of missing speech sounds," *Science*, vol. 167, no. 3917, pp. 392–393, 1970.

[180] T. D. Wager, "Expectations and anxiety as mediators of placebo effects in pain," 2005.

[181] K. Persaud and P. Hemmer, "The influence of knowledge and expectations for color on episodic memory," in *Proceedings of the annual meeting of the cognitive science society*, vol. 36, 2014.

[182] K. J. Klaaren, S. D. Hodges, and T. D. Wilson, "The role of affective expectations in subjective experience and decision-making," *Social Cognition*, vol. 12, no. 2, pp. 77–101, 1994.

[183] S. J. Boyce, A. Pollatsek, and K. Rayner, "Effect of background information on object identification.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 15, no. 3, p. 556, 1989.

[184] N. Gekas, A. R. Seitz, and P. Seriès, "Expectations developed over multiple timescales facilitate visual search performance," *Journal of vision*, vol. 15, no. 9, pp. 10–10, 2015.

[185] L. Elazary and L. Itti, "A bayesian model for efficient visual search and recognition," *Vision research*, vol. 50, no. 14, pp. 1338–1352, 2010.

[186] O. Shurygina, Á. Kristjánsson, L. Tudge, and A. Chetverikov, "Expectations and perceptual priming in a visual search task: Evidence from eye movements and behavior.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 45, no. 4, p. 489, 2019.

[187] G. Buckingham, J. Cant, and M. Goodale, "Living in a material world: How visual cues to material properties affect the way that we lift objects and perceive their weight," *Journal of neurophysiology*, vol. 102, pp. 3111–8, 09 2009.

[188] R. Fleming, C. Wiebel, and K. Gegenfurtner, "Perceptual qualities and material classes," *Journal of vision*, vol. 13, 07 2013.

[189] L. Alley, A. Schmid, and K. Doerschner, "Visual perception of surprising materials in dynamic scenes," *BioRxiv*, 08 2019.

[190] D. Kersten, P. Mamassian, and A. Yuille, "Object perception as bayesian inference," *Annual review of psychology*, vol. 55, pp. 271–304, 02 2004.

[191] K. Kveraga, G. Avniel, and M. Bar, "Top-down predictions in the cognitive brain," *Brain and cognition*, vol. 65, pp. 145–68, 12 2007.

[192] C. J. Downing, "Expectancy and visual-spatial attention: effects on perceptual quality.," *Journal of Experimental Psychology: Human perception and performance*, vol. 14, no. 2, p. 188, 1988.

[193] F. P. de Lange, M. Heilbron, and P. Kok, "How Do Expectations Shape Perception?," *Trends in Cognitive Sciences*, vol. 22, no. 9, pp. 764–779, 2018.

[194] B. M. Urgen and H. Boyaci, "Unmet expectations delay sensory processes," *Vision Research*, vol. 181, pp. 1–9, 2021.

[195] E. De Loof, F. Van Opstal, and T. Verguts, "Predictive information speeds up visual awareness in an individuation task by modulating threshold setting, not processing efficiency," *Vision research*, vol. 121, pp. 104–112, 2016.

[196] P. Kok, J. F. Jehee, and F. P. De Lange, "Less is more: expectation sharpens representations in the primary visual cortex," *Neuron*, vol. 75, no. 2, pp. 265–270, 2012.

[197] T. Stein and M. V. Peelen, "Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity.," *Journal of Experimental Psychology: General*, vol. 144, no. 6, p. 1089, 2015.

[198] A. Y.-C. Chang, R. Kanai, and A. K. Seth, "Cross-modal prediction changes the timing of conscious access during the motion-induced blindness," *Consciousness and cognition*, vol. 31, pp. 139–147, 2015.

[199] Y. Pinto, S. van Gaal, F. P. de Lange, V. A. Lamme, and A. K. Seth, "Expectations accelerate entry of visual stimuli into awareness," *Journal of Vision*, vol. 15, no. 8, pp. 13–13, 2015.

[200] M. Chalk, A. R. Seitz, and P. Seriès, "Rapidly learned stimulus expectations alter perception of motion," *Journal of vision*, vol. 10, no. 8, pp. 2–2, 2010.

[201] P. Sterzer, C. Frith, and P. Petrovic, "Believing is seeing: expectations alter visual awareness," *Current Biology*, vol. 18, no. 16, pp. R697–R698, 2008.

[202] P. Mamassian and R. Goutcher, "Prior knowledge on the illumination position," *Cognition*, vol. 81, no. 1, pp. B1–B9, 2001.

[203] Y. Weiss, E. P. Simoncelli, and E. H. Adelson, "Motion illusions as optimal percepts," *Nature neuroscience*, vol. 5, no. 6, pp. 598–604, 2002.

[204] R. Becklen and D. Cervone, "Selective looking and the noticing of unexpected events," *Memory & cognition*, vol. 11, pp. 601–608, 1983.

[205] A. Mack, "Inattentional blindness: Looking without seeing," *Current directions in psychological science*, vol. 12, no. 5, pp. 180–184, 2003.

[206] D. J. Simons and C. F. Chabris, "Gorillas in our midst: Sustained inattentional blindness for dynamic events," *perception*, vol. 28, no. 9, pp. 1059–1074, 1999.

[207] P. Seriès and A. Seitz, "Learning what to expect (in visual perception)," *Frontiers in human neuroscience*, vol. 7, p. 668, 2013.

[208] W. J. Adams, E. W. Graf, and M. O. Ernst, "Experience can change the'light-from-above'prior," *Nature neuroscience*, vol. 7, no. 10, pp. 1057–1058, 2004.

[209] J. Sun and P. Perona, "Where is the sun?," *Nature neuroscience*, vol. 1, no. 3, pp. 183–184, 1998.

[210] W. Hershberger, "Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below.," *Journal of comparative and physiological psychology*, vol. 73, no. 3, p. 407, 1970.

[211] G. Kanizsa, "Convexity and symmetry in figure-ground orgnization," *Vision and artifact*, pp. 25–32, 1976.

[212] M. A. Peterson and E. Salvagio, "Inhibitory competition in figure-ground perception: Context and convexity," *Journal of Vision*, vol. 8, no. 16, pp. 4–4, 2008.

[213] P. C. Quinn and R. S. Bhatt, "Size and orientation cue figure-ground segregation in infants," *Visual cognition*, vol. 26, no. 7, pp. 518–529, 2018.

[214] P. C. Quinn and R. S. Bhatt, "Visual pop-out in young infants: Convergent evidence and an extension," *Infant Behavior and Development*, vol. 21, no. 2, pp. 273–288, 1998.

[215] M. W. Guy, G. D. Reynolds, S. M. Mosteller, and K. C. Dixon, "The effects of stimulus symmetry on hierarchical processing in infancy," *Developmental Psychobiology*, vol. 59, no. 3, pp. 279–290, 2017.

[216] R. S. Bhatt, A. Hayden, and P. C. Quinn, "Perceptual organization based on common region in infancy," *Infancy*, vol. 12, no. 2, pp. 147–168, 2007.

[217] P. Brugger and S. Brugger, "The easter bunny in october: Is it disguised as a duck?," *Perceptual and motor skills*, vol. 76, no. 2, pp. 577–578, 1993.

[218] I. S. Kerrigan and W. J. Adams, "Learning different light prior distributions for different contexts," *Cognition*, vol. 127, no. 1, pp. 99–104, 2013.

[219] A. R. Girshick, M. S. Landy, and E. P. Simoncelli, "Cardinal rules: visual orientation perception reflects knowledge of environmental statistics," *Nature neuroscience*, vol. 14, no. 7, pp. 926–932, 2011.

[220] R. Zhang, O.-S. Kwon, and D. Tadin, "Illusory movement of stationary stimuli in the visual periphery: Evidence for a strong centrifugal prior in motion processing," *Journal of Neuroscience*, vol. 33, no. 10, pp. 4415–4423, 2013.

[221] D. C. Knill and A. Pouget, "The bayesian brain: the role of uncertainty in neural coding and computation," *TRENDS in Neurosciences*, vol. 27, no. 12, pp. 712–719, 2004.

[222] N. Rungratsameetaweemana and J. T. Serences, "Dissociating the impact of attention and expectation on early sensory processing," *Current opinion in psychology*, vol. 29, pp. 181–186, 2019.

[223] P. Hewson, "Bayesian data analysis 3rd edn a. gelman, jb carlin, hs stern, db dunson, a. vehtari and db rubin, 2013 boca raton, chapman and hall–crc 676 pp.,£ 44.99 isbn 1-439-84095-4," *Journal of the Royal Statistical Society Series A*, vol. 178, no. 1, pp. 301–301, 2015.

[224] W. J. Ma, K. Kording, and D. Goldreich, "Bayesian models of perception and action," 2022.

[225] I. Vilares and K. Kording, "Bayesian models: the structure of the world, uncertainty, behavior, and the brain," *Annals of the New York Academy of Sciences*, vol. 1224, no. 1, pp. 22–39, 2011.

[226] T. E. Hudson, L. T. Maloney, and M. S. Landy, "Movement planning with probabilistic target information," *Journal of neurophysiology*, vol. 98, no. 5, pp. 3034–3046, 2007.

[227] W. S. Geisler, J. S. Perry, B. Super, and D. Gallogly, "Edge co-occurrence in natural images predicts contour grouping performance," *Vision research*, vol. 41, no. 6, pp. 711–724, 2001.

[228] T. L. Griffiths and J. B. Tenenbaum, "Optimal predictions in everyday cognition," *Psychological science*, vol. 17, no. 9, pp. 767–773, 2006.

[229] M. Colombo and P. Seriés, "Bayes in the brain—on bayesian modelling in neuroscience. 63 (3), 697-723," 2012.

[230] J. S. Bowers and C. J. Davis, "Bayesian just-so stories in psychology and neuroscience.," *Psychological bulletin*, vol. 138, no. 3, p. 389, 2012.

[231] M. Bar and E. Aminoff, "Cortical analysis of visual context," *Neuron*, vol. 38, no. 2, pp. 347–358, 2003.

[232] M. Bar, R. B. Tootell, D. L. Schacter, D. N. Greve, B. Fischl, J. D. Mendola, B. R. Rosen, and A. M. Dale, "Cortical mechanisms specific to explicit visual object recognition," *Neuron*, vol. 29, no. 2, pp. 529–535, 2001.

[233] M. Bar, K. S. Kassam, A. S. Ghuman, J. Boshyan, A. M. Schmid, A. M. Dale, M. S. Hämäläinen, K. Marinkovic, D. L. Schacter, B. R. Rosen, *et al.*,

"Top-down facilitation of visual recognition," *Proceedings of the national academy of sciences*, vol. 103, no. 2, pp. 449–454, 2006.

[234] M. J. Fenske, E. Aminoff, N. Gronau, and M. Bar, "Top-down facilitation of visual object recognition: object-based and context-based contributions," *Progress in brain research*, vol. 155, pp. 3–21, 2006.

[235] F. P. de Lange, D. A. Rahnev, T. H. Donner, and H. Lau, "Prestimulus oscillatory activity over motor cortex reflects perceptual expectations," *Journal of Neuroscience*, vol. 33, no. 4, pp. 1400–1410, 2013.

[236] J. Fiser, P. Berkes, G. Orbán, and M. Lengyel, "Statistically optimal perception and learning: from behavior to neural representations," *Trends in cognitive sciences*, vol. 14, no. 3, pp. 119–130, 2010.

[237] P. Berkes, G. Orbán, M. Lengyel, and J. Fiser, "Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment," *Science*, vol. 331, no. 6013, pp. 83–87, 2011.

[238] M. L. Platt and P. W. Glimcher, "Neural correlates of decision variables in parietal cortex," *Nature*, vol. 400, no. 6741, pp. 233–238, 1999.

[239] M. C. Dorris and D. P. Munoz, "Saccadic probability influences motor preparation signals and time to saccadic initiation," *Journal of Neuroscience*, vol. 18, no. 17, pp. 7015–7026, 1998.

[240] M. A. Basso and R. H. Wurtz, "Modulation of neuronal activity by target uncertainty," *Nature*, vol. 389, no. 6646, pp. 66–69, 1997.

[241] A. Todorovic, F. van Ede, E. Maris, and F. P. de Lange, "Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an meg study," *Journal of Neuroscience*, vol. 31, no. 25, pp. 9118–9123, 2011.

[242] T. Yoshiura, J. Zhong, D. K. Shibata, W. E. Kwok, D. A. Shrier, and Y. Numaguchi, "Functional mri study of auditory and visual oddball tasks," *Neuroreport*, vol. 10, no. 8, pp. 1683–1688, 1999.

[243] R. Marois, H.-C. Leung, and J. C. Gore, "A stimulus-driven approach to object identity and location processing in the human brain," *Neuron*, vol. 25, no. 3, pp. 717–728, 2000.

[244] T. Baldeweg, "Erp repetition effects and mismatch negativity generation: a predictive coding perspective," *Journal of Psychophysiology*, vol. 21, no. 3-4, pp. 204–213, 2007.

[245] C. Haenschel, D. J. Vernon, P. Dwivedi, J. H. Gruzelier, and T. Baldeweg, "Event-related brain potential correlates of human auditory sensory memory-trace formation," *Journal of Neuroscience*, vol. 25, no. 45, pp. 10494–10501, 2005.

[246] D. A. Kaliukhovich and R. Vogels, "Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex," *Cerebral cortex*, vol. 21, no. 7, pp. 1547–1558, 2011.

[247] J. S. Isaacson and M. Scanziani, "How inhibition shapes cortical activity," *Neuron*, vol. 72, no. 2, pp. 231–243, 2011.

[248] B. M. Urgen and H. Boyaci, "A recurrent cortical model can parsimoniously explain the effect of expectations on sensory processes," *bioArxiv*, 2021.

[249] R. W. Fleming, K. R. Gegenfurtner, and S. Nishida, "Visual perception of materials: the science of stuff.," *Vision research*, vol. 109, pp. 123–124, 2015.

[250] W. Fujisaki, N. Goda, I. Motoyoshi, H. Komatsu, and S. Nishida, "Audiovisual integration in the human perception of materials," *Journal of Vision*, vol. 14, no. 4, pp. 12–12, 2014.

[251] E. H. Adelson, "On seeing stuff: the perception of materials by humans and machines," in *Human vision and electronic imaging VI*, vol. 4299, pp. 1–12, SPIE, 2001.

[252] L. Sharan, R. Rosenholtz, and E. H. Adelson, "Accuracy and speed of material categorization in real-world images," *Journal of vision*, vol. 14, no. 9, pp. 12–12, 2014.

[253] C. B. Wiebel, M. Valsecchi, and K. R. Gegenfurtner, "The speed and accuracy of material recognition in natural images," *Attention, Perception, & Psychophysics*, vol. 75, pp. 954–966, 2013.

[254] S. C. Pont and J. J. Koenderink, "Bidirectional texture contrast function," *International Journal of Computer Vision*, vol. 62, pp. 17–34, 2005.

[255] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 11, pp. 2032–2047, 2008.

[256] L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson, "Recognizing materials using perceptually inspired features," *International journal of computer vision*, vol. 103, pp. 348–371, 2013.

[257] R. W. Fleming, C. Wiebel, and K. Gegenfurtner, "Perceptual qualities and material classes," *Journal of vision*, vol. 13, no. 8, pp. 9–9, 2013.

[258] G. D. S. Ludden, H. N. J. Schifferstein, and P. Hekkert, "Surprise as a Design Strategy," *Design Issues*, vol. 24, no. 2, pp. 28–38, 2008.

[259] G. D. S. Ludden, H. N. J. Schifferstein, and P. Hekkert, "Visual–Tactual Incongruities in Products as Sources of Surprise," *Empirical Studies of the Arts*, vol. 27, no. 1, pp. 61–87, 2009.

[260] T. Stein and M. Peelen, "Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity," *Journal of experimental psychology. General*, vol. 144, 10 2015.

[261] V. Wyart, A. Nobre, and C. Summerfield, "Dissociable prior influences of signal probability and relevance on visual contrast sensitivity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 3593–8, 02 2012.

[262] C. Summerfield and F. de Lange, "Expectation in perceptual decision making: Neural and computational mechanisms," *Nature reviews. Neuroscience*, vol. 15, 10 2014.

[263] R. W. Fleming, "Visual perception of materials and their properties," *Vision research*, vol. 94, pp. 62–75, 2014.

[264] B. L. Anderson, "Visual perception of materials and surfaces," *Current biology*, vol. 21, no. 24, pp. R978–R983, 2011.

[265] A. C. Schmid and K. Doerschner, "Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials," *Journal of vision*, vol. 18, no. 1, pp. 14–14, 2018.

[266] D. Burr and J. Ross, "Visual processing of motion," *Trends in Neurosciences*, vol. 9, pp. 304–307, 1986.

[267] D. H. Brainard, "The Psychophysics Toolbox," *Spatial Vision*, vol. 10, pp. 433–436, 1997.

[268] V. Paulun, F. Schmidt, J. Assen, and R. Fleming, "Shape, motion, and optical cues to stiffness of elastic objects," *Journal of Vision*, vol. 17(1), pp. 1–22, 01 2017.

[269] F. Schmidt, V. Paulun, J. Assen, and R. Fleming, "Inferring the stiffness of unfamiliar objects from optical, shape, and motion cues," *Journal of Vision*, vol. 17(3), pp. 1–17, 03 2017.

[270] K. Doerschner, R. W. Fleming, O. Yilmaz, P. R. Schrater, B. Hartung, and D. Kersten, "Visual motion and the perception of surface material," *Current Biology*, vol. 21, no. 23, pp. 2010–2016, 2011.

[271] C. Sutter, K. Drewing, and J. Müsseler, "Multisensory integration in action control," 2014.

[272] D. Yon, F. P. de Lange, and C. Press, "The Predictive Brain as a Stubborn Scientist," *Trends in Cognitive Sciences*, vol. 23, no. 1, pp. 6–8, 2019.

[273] G. Sotiropoulos, A. R. Seitz, and P. Seriès, "Changing expectations about speed alters perceived motion direction," *Current Biology*, vol. 21, no. 21, pp. R883–R884, 2011.

[274] W. J. Adams, I. S. Kerrigan, and E. W. Graf, "Efficient visual recalibration from either visual or haptic feedback: the importance of being wrong," *Journal of Neuroscience*, vol. 30, no. 44, pp. 14745–14749, 2010.

[275] B. Urgen and H. Boyaci, "When expectations are not met: unraveling the computational mechanisms underlying the effect of expectation on perceptual thresholds," *BioRxiv*, 02 2019.

[276] S. Bitzer, H. Park, F. Blankenburg, and S. J. Kiebel, "Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model.," *Frontiers in human neuroscience*, vol. 8, no. February, pp. 102:1–17, 2014.

[277] D. J. Heeger, "Theory of cortical function," *Proceedings of the National Academy of Sciences*, vol. 114, pp. 1773–1782, feb 2017.

[278] K. Friston, "The history of the future of the bayesian brain," *NeuroImage*, vol. 62, no. 2, pp. 1230–1233, 2012.

[279] D. C. Knill, "Robust cue integration: A bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant," *Journal of vision*, vol. 7, no. 7, pp. 5–5, 2007.

[280] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.

[281] A. Ferrari, D. Richter, and F. P. de Lange, "Updating contextual sensory expectations for adaptive behavior," *Journal of Neuroscience*, vol. 42, no. 47, pp. 8855–8869, 2022.

[282] C. Song and G. Rees, "Intra-hemispheric integration underlies perception of tilt illusion," *Neuroimage*, vol. 175, pp. 80–90, 2018.

[283] S. O. Murray, H. Boyaci, and D. Kersten, "The representation of perceived angular size in human primary visual cortex," *Nature neuroscience*, vol. 9, no. 3, pp. 429–434, 2006.

[284] I. Sperandio and P. A. Chouinard, "The mechanisms of size constancy," *Multisensory research*, vol. 28, no. 3-4, pp. 253–283, 2015.

[285] B. Zenger-Landolt and D. J. Heeger, "Response suppression in v1 agrees with psychophysics of surround masking," *Journal of Neuroscience*, vol. 23, no. 17, pp. 6884–6893, 2003.

[286] F. Fang, H. Boyaci, D. Kersten, and S. O. Murray, "Attention-dependent representation of a size illusion in human v1," *Current biology*, vol. 18, no. 21, pp. 1707–1712, 2008.

[287] S. Treue, "Perceptual enhancement of contrast by attention," *Trends in cognitive sciences*, vol. 8, no. 10, pp. 435–437, 2004.

[288] U. T. Peter, "Voluntary attention modulates the brightness of overlapping transparent surfaces," *Vision research*, vol. 45, no. 9, pp. 1095–1098, 2005.

[289] A.-M. Bonnel, J.-F. Stein, and P. Bertucci, "Does attention modulate the perception of luminance changes?," *The Quarterly Journal of Experimental Psychology Section A*, vol. 44, no. 4, pp. 601–626, 1992.

[290] W. Prinzmetal, I. Nwachuku, L. Bodanski, L. Blumenfeld, and N. Shimizu, "The phenomenology of attention," *Consciousness and cognition*, vol. 6, no. 2-3, pp. 372–412, 1997.

[291] D. T. Levin and M. R. Banaji, "Distortions in the perceived lightness of faces: the role of race categories.," *Journal of Experimental Psychology: General*, vol. 135, no. 4, p. 501, 2006.

[292] C. Firestone and B. J. Scholl, "Can you experience 'top-down'effects on perception?: The case of race categories and perceived lightness," *Psychonomic bulletin & review*, vol. 22, pp. 694–700, 2015.

[293] L. Yan, Y. Zhu, Y. Shen, Y. Liang, Z. Wang, Y.-H. P. Sun, and N. G. Xiao, "Race categories modulated the perceived lightness of faces," *Available at SSRN 4108571*, 2022.

[294] D. H. Chang, Y. Y. Cheang, and M. So, "Contextual effects in face lightness perception are not expertise-dependent," *Vision*, vol. 2, no. 2, p. 23, 2018.

[295] Z. Pylyshyn, "Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception," *Behavioral and brain sciences*, vol. 22, no. 3, pp. 341–365, 1999.