At which point does noise masking cause a layer scission?

Hannah Louisa Boldt h.boldt@campus.tu-berlin.de
410057

December 20, 2024

Bachelor Thesis Computer Science (B.Sc.)

Technische Universität Berlin, Berlin, Germany
Fakulty IV: Electrical Engineering and Computer Science,
Institute of Computer Engineering and Microelectronics,
Dept. Computational Psychology

First examiner: Prof. Dr. Marianne Maertens
Second examiner: Prof. Dr. Guillermo Gallego
Supervisor: Dr. Guillermo Aguilar

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit eigenständig ohne Hilfe Dritter und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Alle Stellen die den benutzten Quellen und Hilfsmitteln unverändert oder sinngemäß entnommen sind, habe ich als solche kenntlich gemacht.

Sofern generative KI-Tools verwendet wurden, habe ich Produktnamen, Hersteller, die jeweils verwendete Softwareversion und die jeweiligen Einsatzzwecke (z.B. sprachliche Überprüfung und Verbesserung der Texte, systematische Recherche) benannt. Ich verantworte die Auswahl, die Übernahme und sämtliche Ergebnisse des von mir verwendeten KI-generierten Outputs vollumfänglich selbst.

Die Satzung zur Sicherung guter wissenschaftlicher Praxis an der TU Berlin vom 15. Februar 2023. https://www.static.tu.berlin/fileadmin/www/10002457/K3-AMBl/Amtsblatt_2023/Amtliches_Mitteilungsblatt_Nr._16_vom_30.05.2023.pdf habe ich zur Kenntnis genommen.

Ich erkläre weiterhin, dass ich die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt habe.

Berlin, den 19. Dezember 2024

Hannah Louisa Boldt .... (Hannah Louisa Boldt)

Zusammenfassung

Bei welchem Punkt initiiert das Maskieren mit Rauschen eine Ebenentrennung?
Das Ziel von Forschern der visuellen Wahrnehmung ist es zu verstehen, wie unser visuelles System funktioniert. Ein Teil davon ist zu verstehen, wie wir und wann wir Bilder in verschiedene Ebenen rekonstruieren. Dies könnte uns helfen zu verstehen, wie wir Helligkeit wahrnehmen. Ebenentrennung in Form von Transparenz ist ein Phänomen, das schwer vorherzusagen ist. Denn es gibt mehrere Reize sowohl auf niedrigem Level als auch auf hohem Level wie zum Beispiel durch Kontext, der beeinflussen kann, ob Transparenz wahrgenommen wird oder nicht. In dieser Bachelorarbeit fokussieren wir uns auf Mechanismen bzw. Reizen auf niedrigem Level in einfachen Stimuli. Wir verwenden Versionen der mit Narrowband-Rauschen maskierten Whites Illusion, um zu testen, bei welchen Versionen Transparenz auftritt. Da unsere Masken die gesamte Whites Illusion überdecken, fehlen offensichtliche Reize wie zum Beispiel ein niedrigerer Kontrast in der Region der Überlappung beider Ebenen. Wir zeigen per Experiment, dass softe X-Kreuzungen in diesem Fall wahrscheinlich notwendig sind, um Transparenz wahrnehmen zu können und dass es vermutlich eine Mindestgröße, welche durch Kontrastsensitivität vermutlich geändert werden könnte, für diese Kreuzungen gibt. Basierend auf unseren Daten sehen Beobachter bei niedrigen Frequenzen der Maske (0,1 bis 2 cpd) sehr überzeugt Transparenz und bei hohen Frequenzen der Maske (9 und 12 cpd) sehr überzeugt keine Transparenz. Beobachter sind sich unsicher bei Frequenzen der getesteten Maske zwischen 2,5 bis 5 cpd. Diese Punkte werden von der Frequenz des Gitters beeinflusst: Je höher die Frequenz des Gitters, desto höher ist die Frequenz der Maske, bei welcher ein Beobachter sich am unsichersten in seiner Entscheidung befindet. Obwohl Frequenz eine wichtige Variable zu sein scheint, wurde diese in vielen Studien zur Wahrnehmung von Transparenz nicht miteinbezogen in den Stimuli.

At which point does noise masking cause a layer scission?

Hannah Louisa Boldt h.boldt@campus.tu-berlin.de
410057

December 20, 2024

Abstract

The main goal that vision researchers have is to understand how our visual system works. Part of that is understanding how the visual system reconstructs an image into multiple layers. This could help us understand how we perceive lightness. Layer scission in the form of transparency is difficult to predict. There are multiple low-level cues and high-level cues, which are derived from context, which influence whether layer scission is perceived or not. In this thesis, we will focus on low-level mechanisms in simplified stimuli. We use versions of noise masked White's illusion to test at which variables transparency occurs. Because our mask overlays the entire White's illusion, obvious cues like a decrease in contrast are absent. We present that soft X-junctions might be the primary cues for initiating transparency and that the size of the junction and contrast sensitivity shift the perception. Based on our data, at low noise frequencies (0.1 to 2 cpd), observers are confident in perceiving transparency, and at high frequencies 9 to 12 cpd are confident in not perceiving transparency. Also, we derive that at noise frequencies from 2.5 to 5 cpd and grating frequencies from 0.2 to 0.8 cpd, observers are not confident in telling whether there is transparency. These points, thresholds, are affected by grating frequency. Although frequency seems to be an important variable, it has been left out in many studies studying human transparency perception. We think that they might help models predicting human transparency more robust.

Contents

Introduction

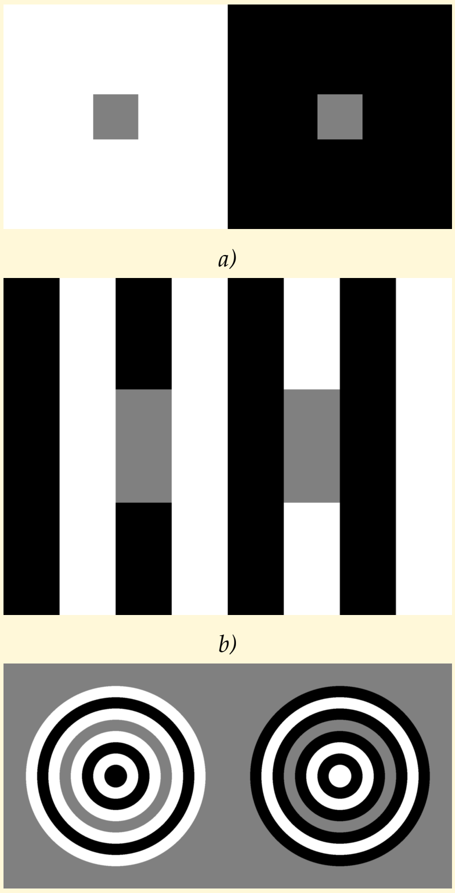
Visual perception is an essential part of our lives. It helps us navigate through life by enabling us to detect objects and avoid objects. We live in a three-dimensional space and have learned to recognize when objects are close to us, when they are far away, or when they are overlapping. Depth perception is also because we have two eyes that see two different images that together create depth. However, even with one eye closed, we can tell whether an object is closer, further away, or in front of another. Similarly, when we see photographs, we can also perceive foreground and background, even when there is no depth in the image. The amount of light our eye reaches is a combination of physical causes, but we do not see light, as a sensor of a camera would; instead, our vision system adapts to the light, and we process that information before we perceive light. This thesis is concerned with when the perception of depth in certain images arises. *To understand and motivate the question: At which point does noise masking cause a layer scission? We will first give an overview of lightness illusions 1) to give context to understand where the noise masked stimuli from the experiment in this thesis came from and 2) to define important terms to characterize images.*

Background on Lightness illusions

Lightness illusions and basic terminology While ***luminance*** is the physical amount of light reflected measured in cd/m^2, ***lightness*** refers to the perceived luminance. [1] ***Lightness*** to refer to the perceived reflectance. The literatur the terms *lightness* and *brightness* are sometimes used interchangably. Lightness is perceived based on the context and not only by the amount of physical light. This can be observed in Figure 1. In each example, the two grey patches have the same luminance. Yet, to most observers, the two equiluminant grey patches are perceived as having different lightness.

Even though visual perception is so essential, the core processes and mechanisms that are at play in visual perception have not been fully understood, and there is no coherent theory that explains how we perceive lightness. Edges play a role in these illusions. An ***edge*** is defined as a discontinuity in luminance. In the *simultaneous contrast* (Figure 1 a)), the patch on the left, which shares edges with the light background, is perceived darker than the patch on the right, which shares edges with the dark background. In *White's illusion* (Figure 1 b)), the patches are embedded on either a dark bar or light bar, and both patches share edges with light and dark regions. Contrary to the simultaneous contrast, the left patch, which shares more edges with the light region, is perceived as lighter. Edges orthogonal to the grating, which may play an important role in this illusion, also exist (Betz et al., 2015b). In the *circular White's illusion* (Figure 1 c)), the left patch only shares edges with a lighter region, similar to the simultaneous contrast; however, the opposite effect is perceived. The same effect as in the White's illusion is perceived, even without edges orthogonal to the grating.

The masked White's illusion used by Betz et al. (2015a) in a slightly altered version will be used in this thesis.



a)

b)

c)

Figure 1: *Recreated versions of a) Simultaneous contrast, b) White's illusion (White, 1979) and c) Circular White's illusion (Howe, 2005). In each stimulus both grey patches are perceived as one being darker or lighter than the other even though they are all equiluminant.*

Noise masks have been applied to lightness illusions and we will give context on the reason behind their usage. Afterward we will explain the theories behind layer scission, specifically transparency.

Importance of edges in lightness illusions There is no coherent theory that explains all three lightness

[1] *Reflectance* describes the ratio of reflected physical light of a surface. Sometimes perceived reflectance is referred to as *lightness* and perceived luminance referred to as *brightness* Betz, 2016).

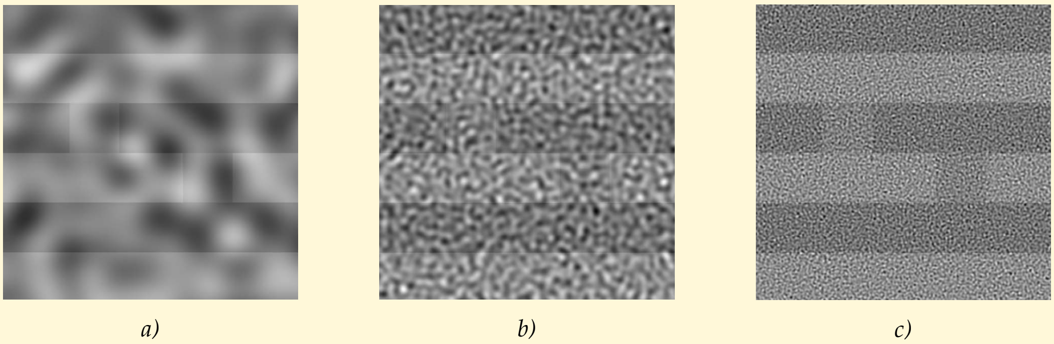


Figure 2: *a) from Betz et al., 2015a a) 0.58 cpd, b) 3cpd c) 9cpd (if viewed on A4 at approximately 17 cm distance) White's illusion with narrowband noise at different frequencies.*

illusions, but edges seem to play a critical role. By masking edges, the effect of illusionary lightness can be increased or reduced. Betz et al., 2015b; Salmela and Laurinen, 2009; Betz et al., 2015a). For example, in White's illusion, edges orthogonal to the grating create junctions, which might then signal our visual system that the patch belongs to the bar it lays on (see Figure 3; Gilchrist et al., 1999). This might cause the edge between the patch and the carrier bar to exert a stronger influence than the edge between the patch and the neighboring bar, which results in an effect similar to the simultaneous contrast, with the carrier bar acting as the background. By masking this critical edge, the illusion strength can be reduced (Salmela and Laurinen, 2009). An illustration of masking edges (in all orientations) with narrowband noise can be seen in Figure 2.

Edges appear to be perceived more easily by subjects than lightness and there is overall evidence that edge perception influences lightness perception (Salmela and Laurinen, 2009; Maertens and Wichmann, n.d.). Edge detection appears to be one of the initial steps for lightness perception.

However, edges are not only important for lightness perception but also feature detection because they often demarcate the boundaries of objects. For example, distinguishing between foreground, background, and transparent objects (Zaidi et al., 1997; Anderson, 1997). While Noise masking experiments show that the noise mask can interfere with the edges of the underlying image, another effect can be perceived: At some noise conditions, the noise mask is seen as a separate layer on top of the image, like fog. At high frequencies, the noise is seen as part of the image itself, similar to a pattern on a carpet. This switch in perceived layers has been noted by participants for stimuli from Betz et al. (2015a) and seems to relate to some mechanism involving edge perception.

Most people see the noise mask as a separate layer in Figure 2 *a), but this effect disappears in* Figure 2 *c). This thesis investigates the point at which this switch happens in noise masked White's illusion. The following section gives some context about layer scission and defines some terminology. The clearly defined aim of this thesis is on page 5.*

Layer scission

A ***layer scission*** is often defined as the perceptual parting of an image into multiple layers. It describes the perception of multiple layers as layers partially or fully laid on top of each other. An Image is two-dimensional, yet, we can perceive layers and depth in an image. For example clouds in front of a mountain, in which clouds act as the foreground and the mountain as a background (see Figure 5, simplified examples in Figure 3 or Figure 2). For simplicity, instead of writing each time that someone perceives a layer scission, we call the phenomenon of perceiving multiple layers itself as a layer scission.

We will first explain layer scission in form of overlapping layers and then go into detail about layer scision in form of transparency, since this thesis is more concerned about layer scission in form of transparency.

Importance of contrast at edges in layer perception Contrast at edges and change of magnitude in such contrast also appear to be important for layer scission perception (Anderson, Barton L. and Singh, Manish, 2002; Metelli, 1974). We call objects that partially or almost fully let light pass through, such as colored glass, colored water, and smoke ***transparent***, and objects that do not let light pass through are called ***opaque***. ***Transmittance*** is the physical property of a transparent object and describes how transmissive it is (i.e., how much light it lets through). A thin layer of clouds has a higher transmittance than a thick layer of clouds because it lets more light through. The ***Reflectance*** of a surface describes the ratio of how much light is reflected. The terms transparency and transmittance are often used interchangeably; we will use *transparency* to refer to a form of layer scission.

Edges cause junctions and junctions can signal layer scission. If multiple edges meet at one point, they create a junction. A T-junction (see Figure 3) can cause a layer scission between two opaque objects, where the top of the T signals the top layer (Zaidi et al., 1997).

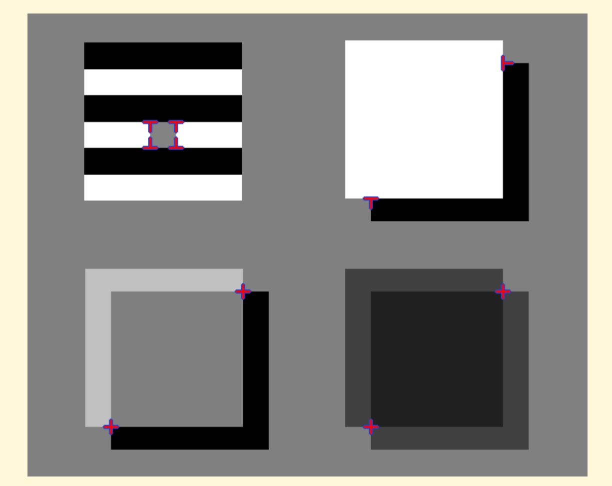


Figure 3: T *and* + *each show* ***T- and X-junctions****, respectively. The first image shows four T-junctions that signal the grouping of the patch to the carrier bar according to Gilchrist et al. (1999). The second image shows two T-junctions that signal a jump between layers, causing the light square to appear in front of the dark square, even though there is no actual depth in this two-dimensional image (similar to (Zaidi et al., 1997).*

The third and fourth images show two X-junctions that can signal a jump between layers of different transparencies. The third shows a transparent light square in front of a black square (Anderson, 1997). In the fourth image, both squares can be seen as transparent objects, and which of them is seen in front is ambiguous. Because of images two and three, the upper square is more likely to be seen in front.

Contrast and contrast polarity Contrast polarity plays an important role in transparency perception. While ***contrast*** describes how big the relational difference in luminance is, ***edge contrast*** refers to the contrast between two regions separated by an edge. ***Contrast polarity*** describes the direction of luminance at one point in some direction (i.e., the direction of contrast on an edge). We define the contrast polarity of an edge from A to B as

$$\text{contrast polarity} := \begin{cases} 1 & \text{if } I_A < I_B, \\ 0 & \text{if } I_A = I_B, \\ -1 & \text{if } I_A > I_B. \end{cases} \tag{1}$$

Where I_A describes the mean luminance *along the edge* of area A and I_B describes mean luminance *along the edge* of area B. Contrast polarity is positive if the step over the edge from A to B is an increment.

Junctions preserving contrast polarity at the edges of the layer below can cause a layer scission in the form of transparency since transparent surfaces only change the degree of the underlying edge contrast, not the contrast polarity at the edge. X-junctions can signal transparency in objects or transparent layers if the contrast polarity remains the same at these X-junctions (Anderson, 1997). For examples see Figure 3 and Figure 4, symbols in pink signify preservation. Therefore, different junctions of edges can determine whether we perceive a layer as opaque or transparent. In natural images, our visual system can easily discern a pane of glass from a landscape behind it, but only if the glass has marks, reflections or filters light.

A cue for layer scission in the form of transparency can be X-junctions, however, it is not necessary to have specifically such junctions in an image to perceive transparent objects. Objects that allow light to pass through, such as smoke, pollen, or wispy clouds, can be easily distinguished from the background. These objects cover the background, but because they let light through, they remain the contrast polarity of the objects behind them (see Figure 5, Figure 4).When both layers are seen, the top layer lowers the contrast of the layer behind. This decrement in the contrast is a cue for transparency. The perceived contrast in the region of overlap determines how transmissive the transparent layer is (Aguilar and Maertens, 2022)

Figure 5: *An example of a cloud varying in opacity covering the edges of the snowy mountains and trees. The contrast polarity at the mountain's contour remains the same.*

Contrast There are multiple definitions for image contrast. One of them is the ***Michelson contrast***. This contrast is calculated by

$$\text{Michelson contrast} := V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \tag{2}$$

Where I_{max} corresponds to the luminance of the light bar and I_{min} corresponds to the luminance of the dark bar. It was used to describe the visibility of a pattern (Michelson, 1995) and therefore is often used in images with simple repeating patterns, such as

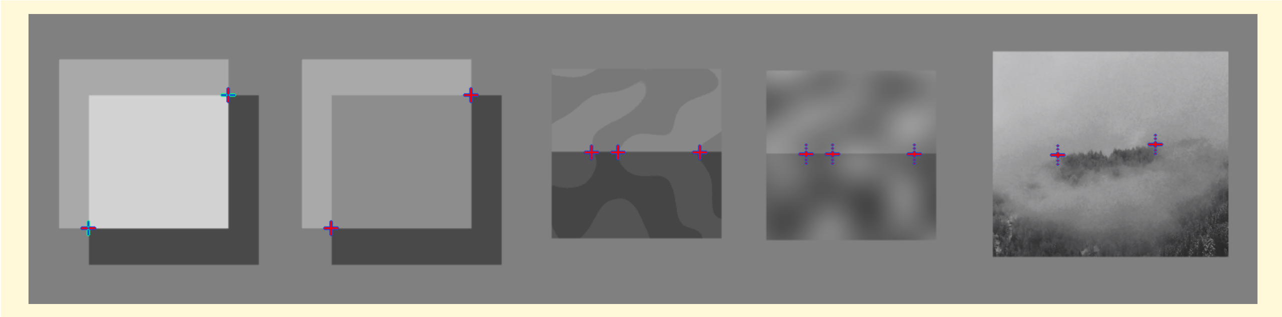


Figure 4: *From left to right: a) No transparency should be observed. Two squares overlap, but the contrast polarity is reversed. See top X-junction: Over the vertical pink contour from left to right, both steps decrease in luminance. But, over the horizontal blue contour, the left step decreases the luminance and the right step increases. Perceptually, the light gray square has reversed the direction of luminance at the edge of the dark gray square beneath. b) A transparent light grey square overlaps a dark square, and contrast polarity is preserved. c) Narrowband noise (like in d)) with thresholding applied such that it only contains two luminance values and two aligned rectangles in varying luminance as the background. Multiple X-junctions at the underlying edge signal transparency. d) Narrowband noise and two aligned rectangles in varying luminance as the background. It contains **Soft X-junctions** at the same places as in c) that signal transparency. Soft refers to the step in luminance over the edge of the narrowband noise mask at the junction with the edge of White's illusion. Compared to c), the luminance step is gradual, soft. e) Clouds in front of the contour or edge of a mountain. Again, Soft X-junctions signal transparency within the clouds.*

a sine grating. Another image contrast is the ***Root Mean Square (RMS) contrast***. This contrast is defined by

$$\text{rms contrast} := \frac{\text{standard deviation}}{\text{mean luminance}}. \tag{3}$$

We have yet to define what a noise mask is or what narrowband noise is. We did this because for that we need to define what a frequency is. And because the frequency can change the contrast sensitivity it is only now that we explain what a narrowband noise mask is.

Patterns across space and noise

Spatial frequency, contrast and noise Spatial ***frequency*** refers to the repetition of a periodic pattern across space and is commonly described in ***cycles per degree (cpd)***. A ***cycle*** refers to one period of a pattern, such as a sine wave on the interval $[0, 2\pi]$ (see Figure 6). One ***degree*** refers to one degree in our field of vision. If a pattern has a frequency of 3 cpd, then this pattern repeats 3 times within one degree.

Figure 6: *Example of a visual pattern with varying cpd. The sinewave-grating starts with low spatial frequency on the left side and logarithmically increases to the right.*

In this thesis, the squarewave-pattern and narrowband noise will be important. A squarewave is composed of bars alternating in luminance and is part of White's illusion.

Contrast sensitivity defined by Campbell and Robson (1968) is as a function that takes in the frequency of a grating and returns the reciprocal of a *threshold contrast* at which the grating is still being perceived or

$$\text{[contrast] sensitivity}(f) := \frac{1}{\text{threshold [contrast]}} \tag{4}$$

where f is the frequency of a pattern or image region and [contrast] is some contrast metric. In Campbell and Robson (1968) the contrast was the Michelson contrast V. So, V sensitivity (f) measures the threshold Michelson contrast of a pattern to be still perceived. This function is collected by showing a pattern at a fixed frequency and then decreasing the contrast until the observer no longer sees the pattern. This contrast level is called the *threshold contrast* for that frequency. This procedure is then repeated for all frequencies to approximate a function based on frequency. Depending on the pattern, the contrast sensitivity function can change. An example of a squarewave grating can be seen in Figure 7.

Noise can be described as taking a sample of values following a random distribution. In images, ***spatial noise*** refers to each pixel following a random distribution. ***spatial white noise*** is, for example, each pixel is independently a fully random value between 0 and 1. An image of spatial noise can be transformed into a so called frequency domain using fast Fourier transform (FFT). Think of it as decomposing an image into multiple sinewave gratings at different frequencies and rotations. Spatial white noise in the frequency domain, therefore, has equally randomly distributed sinewave gratings at different frequencies and rotations. Spatial ***narrowband noise*** is spatial white noise where specific frequencies have been filtered out in the frequency domain. We say a *Narrowband noise*

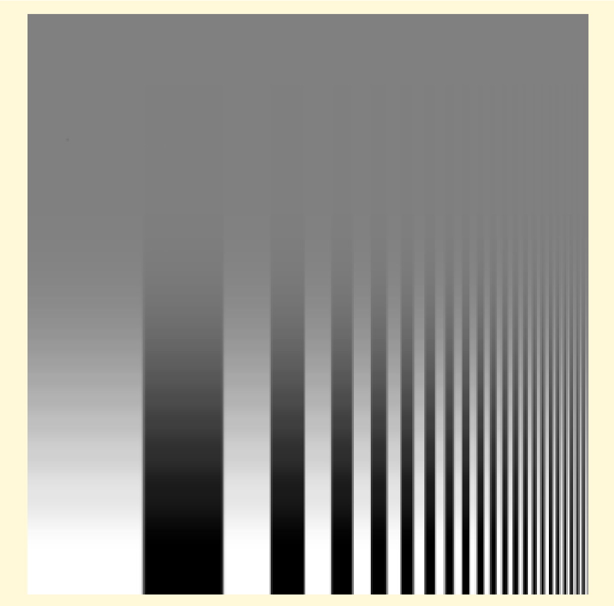


Figure 7: *Example of a squarewave grating and varying contrast at each row. On the x-axis, grating frequency increases logarithmically. On the y-axis, contrast decreases logarithmically. Move further away from the monitor, at some point, the grating is visible in the top mid but not at the top right. The line that separate the visible and invisible region is called the sensitivity function.*

has a frequency of 3 when we have white noise with all frequencies removed but a narrow band around frequency 3. The *bandwidth* describes how wide the band is. All narrowband noise masks in this experiment have a bandwidth of 1, meaning that they contain sinewave gratings of frequency 2 to 4 in all rotations. Examples of narrowband noise at different frequencies can be seen in the methods section (see Figure 9) or overlaid on White's illusion (see Figure 2).

After defining important terms, we will clearly define the aim of this thesis along with the methods to tackle our data question.

Aim of this thesis

We call laying spatial noise on top of existing images ***noise masking***. Noise masking can significantly influence the perception of the image below the mask. Edges, which can be crucial in lightness and layer perception, can be masked by spatial noise. It is argued that when edges are masked with noise, the perception of these edges is reduced. For example, in Sørensen (2023), participants had to draw contours of objects in natural images. Sørensen (2023) demonstrated that noise around 3 cpd disrupts the perception of edges the most, even resulting in contours that were not present in the original image. Therefore, noise masks have been used to mask edges that influence lightness (Figure 2; Salmela and Laurinen, 2009). It has been shown that masking the simultaneous contrast and White's illusion at different frequencies of narrowband noise changed the perceived lightness (Figure 2; Salmela and Laurinen, 2009). This effect can be observed in (Figure 2) and changed depending on the distance to the monitor on which this image is displayed. As we move further away, the image occupies less space in our field of view (i.e., the size of the image measured in degrees decreases, but the number of repetitions stays the same), changing the frequency in cycle per degree.

After noise masking, the mask itself overlaid on top of an image can sometimes be perceived as a transparent layer on top of the image below. In the case of narrowband noise, it can appear as fog or clouds in front of an image. Because masks are applied at a 50% opacity, they mimic transparent objects' characteristics, such as preserving contrast polarity at the edges in the image below. This effect is only sometimes visible to observers and, at high frequencies, seems to disappear. Under which circumstances this effect arises has not been well researched and is the aim of this thesis. We want to determine at which noise frequency our perception of the stimuli switches from a Transparent noise mask on top of White's illusion (yes layer scission) to White's illusion with grainy texture (no layer scission). See Figure 18 and Figure 19 for more examples of noisy images with no layer scission. Based on observations from Betz et al. (2015a) and personal observations before our experiment, we hypothesized that this switch is independent of grating frequency and grating phase. The reason was that the tested grating frequency and grating phase seemed not to change the image content a lot. However, as we will find out, grating frequency did influence the point at which the switch occurred, and while the contrast of the image stayed the same at every condition, the contrast sensitivity might have been influenced by that, causing a shift).

At which point does layer scission in the form of transparency occur in masked White's illusion?

The influence of the frequency of the narrowband noise mask, the frequency of the square grating in White's illusion, and the phase of the square grating in White's illusion

↓

on layer scission in form of transparency compared to the illusion strength of White's illusion.

We tested this by letting observers label each presented image as *yes layer scission* or *no layer scission*.

The following section will explain which methods we used to collect data to answer our data question. We will also describe the laboratory environment and stimuli in detail to give more context to our methods.

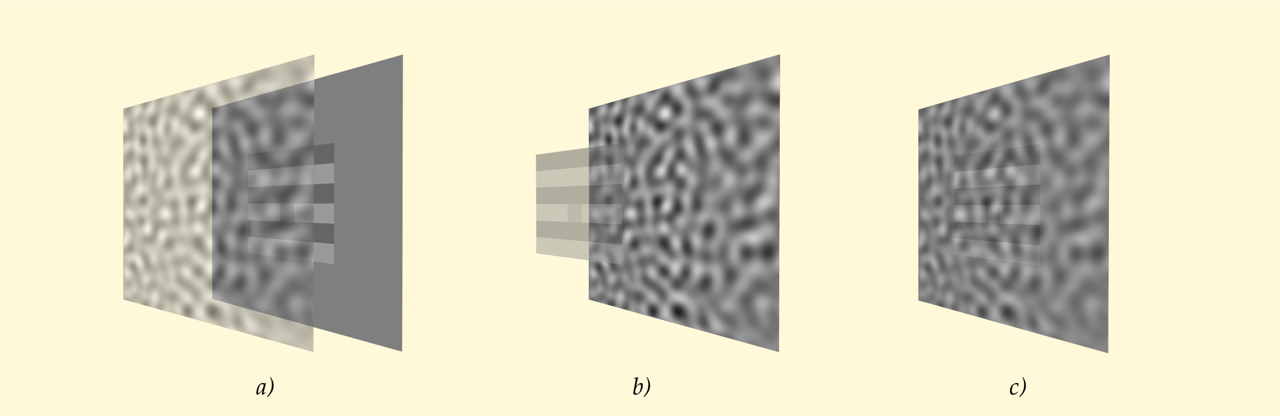


Figure 8: *Illustration of perceptual categorization of the stimulus into multiple layers. a): The stimulus is perceived as two layers. The noise is a transparent layer in front of White's illusion. b) The stimulus is perceived as two layers. White's illusion is a transparent plane in front of the noise, similar to looking through a colored pattern window at clouds. c) The stimulus is perceived as one layer. The noise is part of White's illusion, similar to objects with texture, such as a carpet with White's illusion as a pattern.*

Methods

To test under which conditions noise masking causes a layer scission, stimuli from Betz et al. (2015a) have been adjusted and 3 new noise frequencies have been added. The stimuli's mean luminance and the noise mask's contrast have not been altered. Instead of 0.05 used by Betz et al. (2015a), the new contrast of 0.068 ensured visibility of the grating in every noise condition. This was an important modification since we wanted to test when layer scission occurs between two layers. At a too-low contrast of the underlying image or too-high contrast of the mask, the task would have been futile since only one layer, the mask, would have been visible. Noise masks of frequency 0.1, 0.25, and 12 *cpd* have been added since pilot studies showed that at 0.1 and 0.25, there was a layer scission, and at 12, there was no layer scission.

We will now outline the experimental design and describe independent and dependent variables.

Experimental design

To answer at which point a noise mask causes a layer scission, we performed an experiment that forced participants to choose between two choices: Yes, I perceive a layer scission, and No, I don't perceive a layer scission.

We tested 9 noise frequencies

$$F_n := \{0.1, 0.25, 0.58, 1.0, 2.0, 2.77, 3.55, 9.0, 12.0\} \quad (5)$$

3 grating frequencies

$$F_g := \{0.2, 0.4, 0.8\} \quad (6)$$

and two grating phases

$$\Phi_g := \{0, 1\}. \quad (7)$$

Therefore, observers were shown

$$|F_n| \times |F_g| \times |\Phi_g| = 9 \times 3 \times 2 = 54 \quad (8)$$

stimuli. Each i-th block b_i showed all 54 variations in a randomized order and there were 10 blocks. A new noise mask was generated for every participant for every block.

In our data question *which point* refers to a point in noise frequency and since we have $|F_g| = 3$ grating frequencies, we will predict three points. We will only make predictions about points in the interval $[0.1, 12]$ because the lowest and highest noise frequencies tested were 0.1 and 12. From informal observation, we expected that participants see a layer scission at low frequencies with high certainty and no layer scission at high frequencies with high certainty. At mid frequencies, we expected the lowest certainty. We call this the point at which certainty is the lowest, the *critical point*. We want to estimate the critical point, and therefore, we opted for a discrimination task for each image: yes layer scission or no layer scission. For example, if for noise frequency $f_n = 3.55$ and grating frequency $f_g = 0.8$, an observer has pressed five times yes and five times no, we derive that the critical point is most likely close to 3.55 for this observer at a grating frequency of 0.8. Through multiple repetitions, we can collect the proportion of yes and no responses of all observers. *We will now continue with describing the logic of how stimuli were created, converted and presented to participants.*

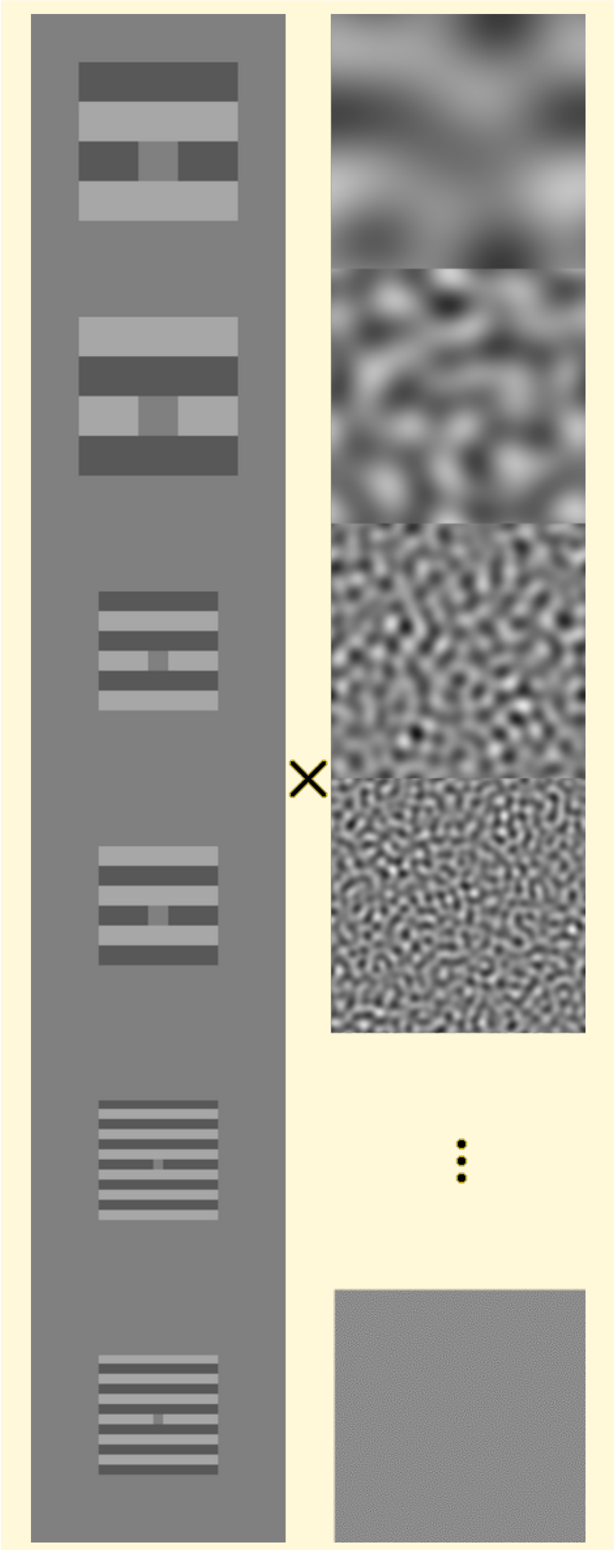


Figure 9: *Shows variations of White's illusion and narrowband noise mask. From top to bottom, frequency increases, and grating phases alternate (luminance of starting bar). Measurements of White's illusion from top to bottom are: size of the illusion* $10.2° \times 10.2°$ *with 4 bars or 0.2 cpd and test patch size* $2.55° \times 2.55°$*, size of the illusion* $7.66° \times 7.66°$ *with 6 bars or 0.4 cpd and test patch size* $1.28° \times 1.28°$*, and size of the illusion* $7.66° \times 7.66°$ *with 12 bars or 0.8 cpd and test patch size* $0.64° \times 0.64°$*. Right side shows variations of the noise mask. Size of the mask used is* $16.3° \times 16.3°$ *and at 0.1, 0.25, 0.58, 1.0, 2.0, 2.77, 3.55, 9.0 and 12.0 cpd. For visualization, the contrast was increased in this illustration.*

Stimuli

This section explains how each stimulus is generated and then converted to be displayed at the desired luminance values. All stimuli were generated in `Python` with `Stimupy` (Schmittwilken et al., 2023) and `Pillow`. Each stimulus is saved as a `dictionary` and contains keys like noise frequency, grating frequency, grating phase and image. The stimulus image itself is stored as a `numpy array` within the dictionary. All stimuli `dictionaries` were then contained in a `numpy array`. This `numpy array` is was then saved as a `.npy` file and loaded at the beginning of the experiment.

First, all the $F_g \times \Phi_g$ pair variations of White's illusion were generated in desired luminance values in cd/m^2. In all White's illusion variations (see Figure 9 left side), the dark bars have a luminance of $41cd/m^2$, the light bars have a luminance of $47cd/m^2$ and the test patches have a luminance of $44cd/m^2$. White's illusion in this experiment has a Michelson contrast of 0.068. The noise masks were narrow band noise at different frequencies but a constant bandwidth of 1 and contrast of 0.2 RMS (standard deviation divided by mean). Each version of White's illusion was then masked by:

$$\texttt{stim["img"]:=} \quad \begin{array}{c} \texttt{0.5 * noise["img"]} \\ + \\ \texttt{0.5 * whites["img"]} \end{array} \qquad (9)$$

This resulted in $10 \times 54 = 540$ stimulus images (see page 9 for all 54 variations).

To display the stimuli at the correct luminance, the desired luminance values were converted to normalized values in [0,1] for the HRL Library to display. For that, 254 normalized values were displayed and luminance was measured by G. Aguilar. The resulting look-up table was then used to fit a linear function with regression (see appendix Figure 24). This function was then inverted to predict normalized values based on luminance. The code for predicting values was done by G. Aguilar. Then, it was modified for faster computation of whole images instead of individual values.

For context, the next section will describe the Lab environment in which the participants were tested and explain the procedure of the experiment in short. The full written instructions for the experiment can be found in figure 25.

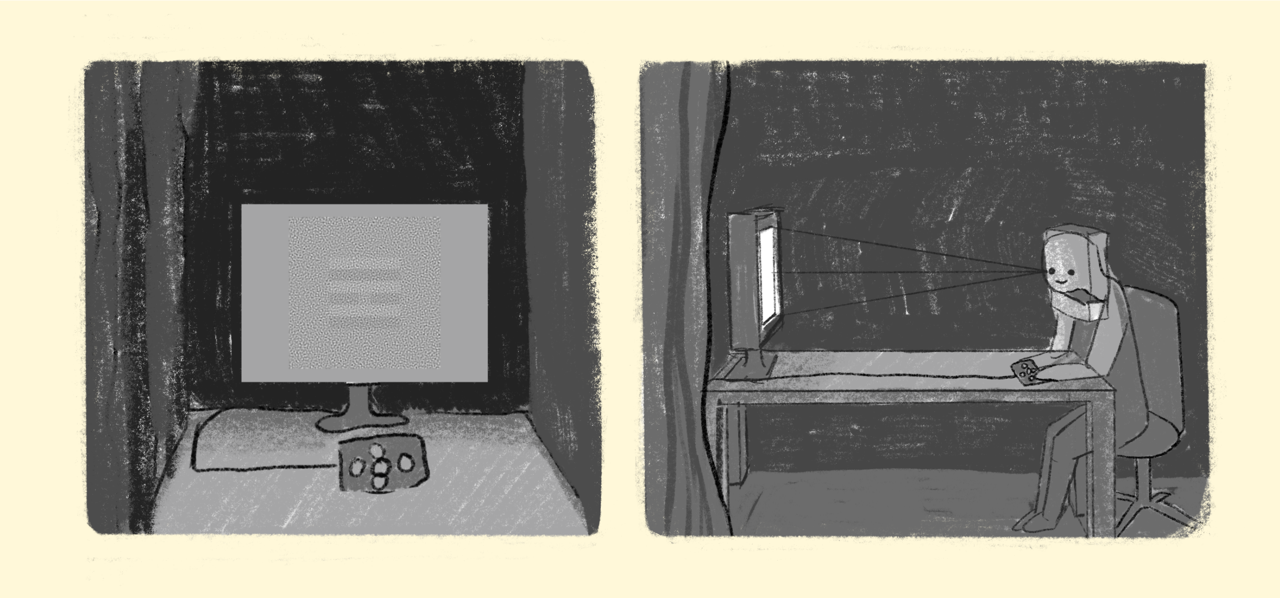


Figure 10: *The left side shows an example view of embedded stimulus on the monitor, and the right side shows the setup with a participant during the experiment.*

Lab Enviroment

Stimuli were displayed on the 120Hz calibrated research-grade display (VIEWPixx / 3D Lite) at 48 pixels per degree (ppd). The monitor is 54 by 29 centimeters and has a 1080 by 720 pixels resolution. At a viewing distance of 75 cm, the resulting stimuli were $16.3° \times 16.3°$. A chinrest enabled a stable viewing distance. To capture the responses from the participants, the RESPONSEPixx 5-button controller has been used (see Figure 10 for set up). To ensure correct constant lighting conditions the monitor had to be turned on 2 hours before the experiment to "warm up" and stay at a constant temperature, since the luminance of the monitor changed with temperature. Additionally, black-out curtains ensured that the light came only from the displayed stimuli and the embedding on the monitor. Before the experiment, all stimuli were generated and saved as dictionaries stored in `NumPy` arrays (Harris et al., 2020). These were then loaded once at the beginning of the experiment. This reduced the run time of the experiment code, which was crucial since there was a fade-in and fade-out time for each stimulus that lasted 6 frames or about 250ms, to reduce after images. Stimuli were displayed using the `HRL` library, provided and developed by `https://github.com/computational-psychology/hrl`.

Procedure

Before the experiment, participants were shown a few images consisting of examples of layer scission to explain the concept in the context of simplified stimuli (see appendix for written instruction). In each trial, a short time (500 *ms*) of viewing the stimuli was set to reduce the strategic reasoning or overthinking in the participants. After the participant saw the stimulus, the screen filled with a gray background at $44cd/m^2$, and the participants then voted whether there was a layer scission. To vote, they pressed the left button on the controller for *No layer scission* and the right button for *Yes layer scission*. After casting a vote, the program continued to the next trial. After every 27th trial, the participants could take a break, but most only took a few breaks. Explaining the concept and conducting the experiment took under an hour. The first 20-30 minutes consisted of doing a letter test to ensure that participants had normal vision, explaining the topic, talking with the participants and asking them some questions. After that, the experiment itself took around 20 minutes (30 minutes with breaks) for most participants.

Participants The experiment was conducted on 6 naive participants and 3 experienced ones (the supervisor, a student writing a thesis on visual perception, and the author). Out of 9 participants, 5 were male and 4 were female. All participants were right-handed except one naive participant.

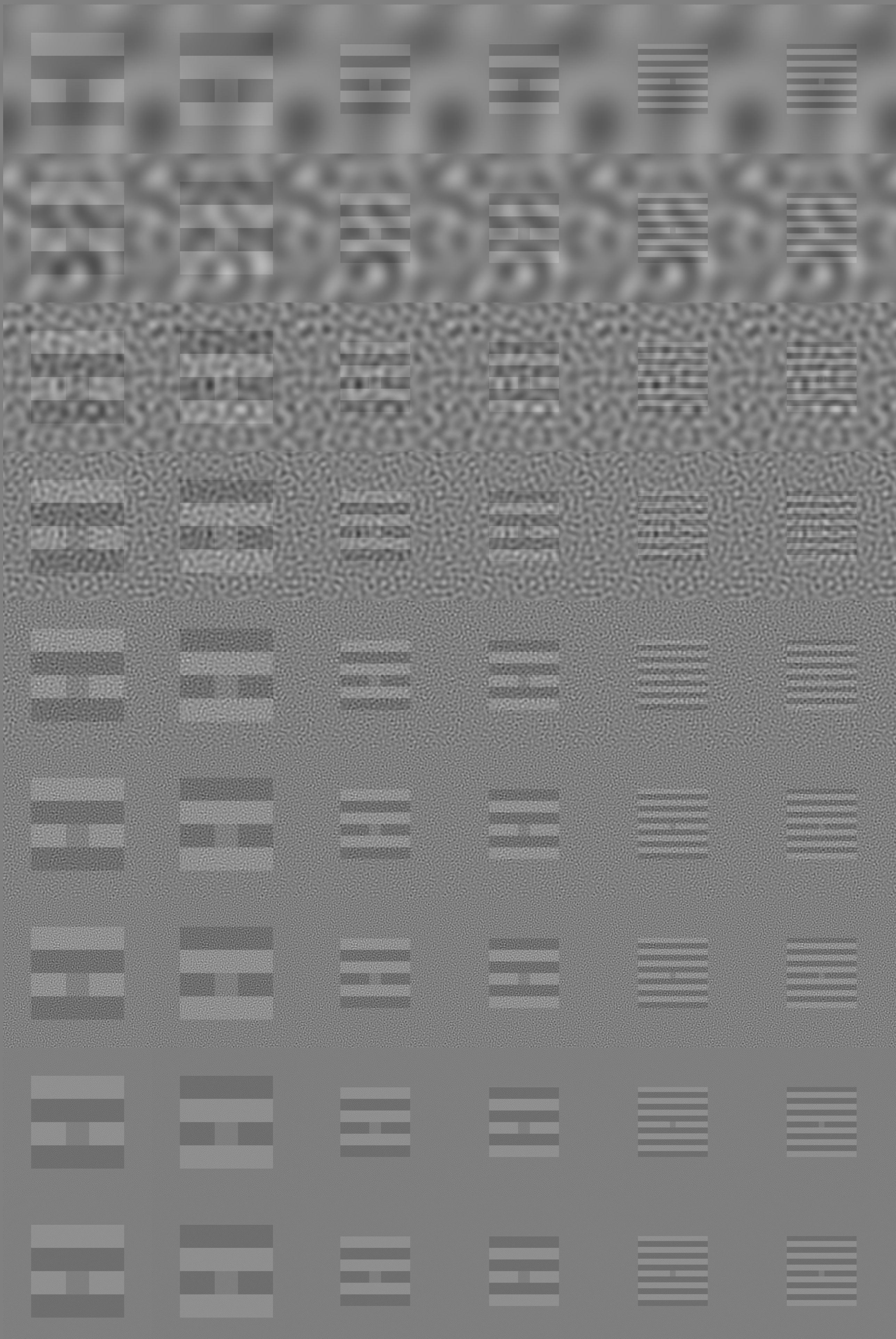


Figure 11: *All 54 Stimuli.*

	Independent Variables			Dependent Variables					Metadata
	Noise frequency	Grating frequency	Grating phase	Layer scission perceived	Response time	Starting time	Drop	Block	Participant
trial: 0	3.55 cpd	0.4 cpd	black	No	5519 ms	2024/11/12/14:42	2 frames	0	A1
trial: 1	0.25 cpd	0.2 cpd	black	Yes	363 ms	2024/11/12/14:42	0 frames	0	A1
trial: 2	2.0 cpd	0.8 cpd	black	Yes	323 ms	2024/11/12/14:42	0 frames	0	A1
trial: 3	9.0 cpd	0.4 cpd	black	No	477 ms	2024/11/12/14:42	0 frames	0	A1
trial: 4	3.55 cpd	0.4 cpd	white	No	272 ms	2024/11/12/14:43	0 frames	0	A1
trial: 5	2.7 cpd	0.8 cpd	white	Yes	336 ms	2024/11/12/14:43	0 frames	0	A1
trial: 6	12.0 cpd	0.4 cpd	black	No	238 ms	2024/11/12/14:43	0 frames	0	A1
trial: 7	0.58 cpd	0.8 cpd	black	Yes	1961 ms	2024/11/12/14:43	0 frames	0	A1
trial: 8	2.7 cpd	0.8 cpd	black	Yes	308 ms	2024/11/12/14:43	4 frames	0	A1
trial: 9	0.1 cpd	0.4 cpd	white	Yes	454 ms	2024/11/12/14:43	0 frames	0	A1

Figure 11: *Raw head of dataframe. Each block contains* 54 *trials and there are* 10 *Blocks numbered from* 0 *to* 9*. This table shows the first ten data points of the first block. There is one table for each participant. Grating phase refers to whether the first bar of white's illusion is light or dark. Drop refers to how many frames have dropped during display. Noise frequency is colored in orange shades and grating frequency is colored in blue shades, darker shades equal higher frequency.*

Results

We found that a layer scission occurs for most observers at low frequencies, and at high noise frequencies, layer scission disappears. For these observers, their data followed a sigmoid-shaped psychometric function, and with increasing grating frequency, the noise frequency for critical points increased. The grating phase did not influence any of the observers.

We will first outline and show the data we collected to give intuition about what data we analyzed. Then, we will introduce the methods used for data analysis. Based on our analysis, we will illustrate the results of this experiment.

We will show three types of functions because they represent the results at different levels: absolute layer scission $LS_{abs\ f_g}$, average layer scission $LS_{avg\ f_g}$ and psychometric layer scission $LS_{psych\ f_g}$. These all measure the impact of noise frequency $f_n \in F_n$ on layer scission at a fixed grating frequency $f_g \in F_g$. Observers varied a lot, simply combining all data from observers and then predicting a point with regression would not have been fruitful, because meaningful differences would have been lost. These functions are independent of the grating phase because the grating phase had little to no influence on whether a layer scission was perceived or not, and therefore, grating phase $\varphi_g \in \Phi_g$ has been counted as another repetition in the following plots and analysis.

Showing raw data

Each participant generated $54 \times 10 = 540$ data points (see Figure 11). One data point contains the noise frequency f_g in cpd, the grating frequency f_n in cpd, the grating phase $_g$ (top bar dark/top bar light), layer scission perceived (yes, no), response time in ms, date and time of the collected data point, drop in frames during the presentation, number of blocks and the alias of the participant.

absolute layer scission We define ***absolute layer scission*** as the response of an observer based on noise frequency, grating frequency, grating phase and block, specifically

$$LS_{abs\ f_g} : F_n \times \Phi_g \times B \to \{0,1\},$$
$$LS_{abs\ f_g}((f_n, \varphi_g, b)) \mapsto \begin{cases} 1, & \text{button press was right} \\ 0, & \text{button press was left.} \end{cases} \tag{10}$$

where again F_n denotes the set of noise frequencies, F_g is the set of frequencies of the grating, Φ_g is the set of the phases of the grating, and B is the set of blocks. Small letters indicate elements of these sets. The following graph shows the impact of individual noise frequencies f_n (x-axis) on the absolute layer scission (y-axis).

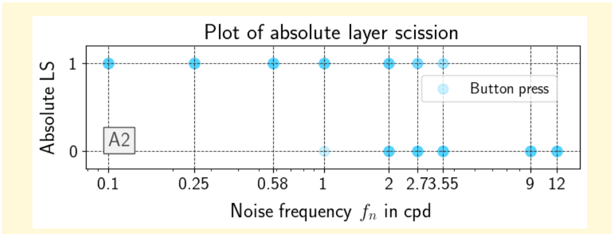


Figure 12: *The x-axis describes the noise mask frequency f_n in cpd, and the y-axis describes $L_{abs\ 0.4}((f_n, b, \varphi_g))$ (i.e., yes (1) or no (0)). Fully colored dots like* (0.58, 1) *correspond to multiple dots overlaid. For each f_n there are $2 \cdot 10 = 20$ dots. This plot only contains data on stimuli with a grating frequency of 0.4 cpd.*

Now we will explain the average layer scission because it estimates a simple approximation of the certainty of layer scission based on grating and noise frequency.

Average Layer scission of participants

Each participant pressed the button 20 times for the same (f_n, f_g) pair which meant that we can define

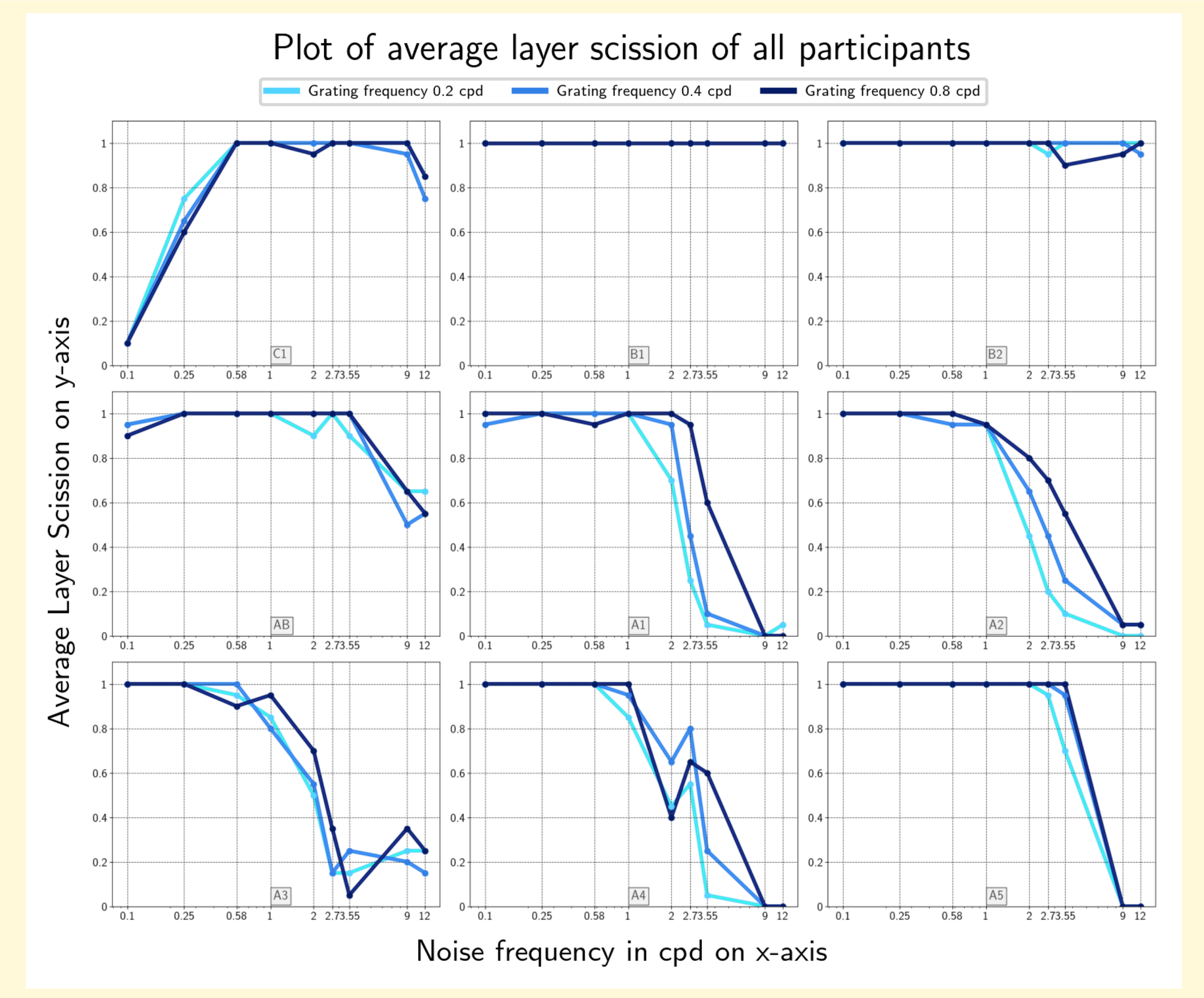


Figure 13: *Average layer scission from all participants. Each graph illustrates the effect of noise frequency f_n and grating frequency f_g on average layer scission perception. The x-axis shows the noise mask center frequency. The y-axis shows the relative layer scission perception. A darker blue shade indicates a higher grating frequency (0.2,0.4,0.8). Grey sticky notes indicate the alias of the observers. In each graph there are $|F_n| \times |F_g| = 9 \times 3 = 27$ blue dots in this plot. Each blue dot is the result of $LS_{avg\ f_g}$ (i.e., mean of $|\varphi_g| \times |b| = 2 \times 10 = 20$ datapoints, grouped by grating and noise frequency). Sometimes, blue dots overlap (top mid: Observer B1 reported a layer scission in every image, so all his averages overlap). Participants A1, ..., A5 were classified as type A observers and B1, B2 as type B observers. Note that each dot in each plot is the result of the average layer scission. The dots were connected for this visualization.*

the ***average layer scission*** as

$$LS_{avg\ f_g} : F_n \to [0,1],$$

$$LS_{avg\ f_g}(f_n) := \frac{1}{20} \sum_{b_i=1}^{10} \sum_{\varphi_g \in \{0,1\}} LS_{abs\ f_g}(f_n, b_i, \varphi_g)). \tag{11}$$

This is the average layer scission of all (f_n, b, φ_g) pairs grouped by noise frequency f_n and grating frequency f_g. It represents an approximation of the certainty of the participant observing a layer scission (see Figure 13). The confidence in a participant's choice is the lowest at an average layer scission of 0.5 because they pressed left 10 times and right 10 times. *When we plot the three average layer scission functions, each varying in shades of blue, we can immediately observe two things: Observers varied a lot and observers who had a switch were affected by the grating phase around the point where their certainty was 0.5.*

Differences between observers *For simplicity, we call the most common type A, the second most common type B, and the remaining participant C1.* Observers of type A classified stimuli with low noise frequency as layer scission and stimuli with high noise frequency as no layer scission. Their classification follows a psychometric function. Observers of Type B classified almost all or all stimuli as layer scission. Their classification follows a constant line at 1. Observer AB was type A for the first five blocks and type B for the last B, but after taking a break. Observer C1 did not observe a layer scission at noise frequency 0.1 but saw a layer scission at noise frequencies beginning from 0.58.

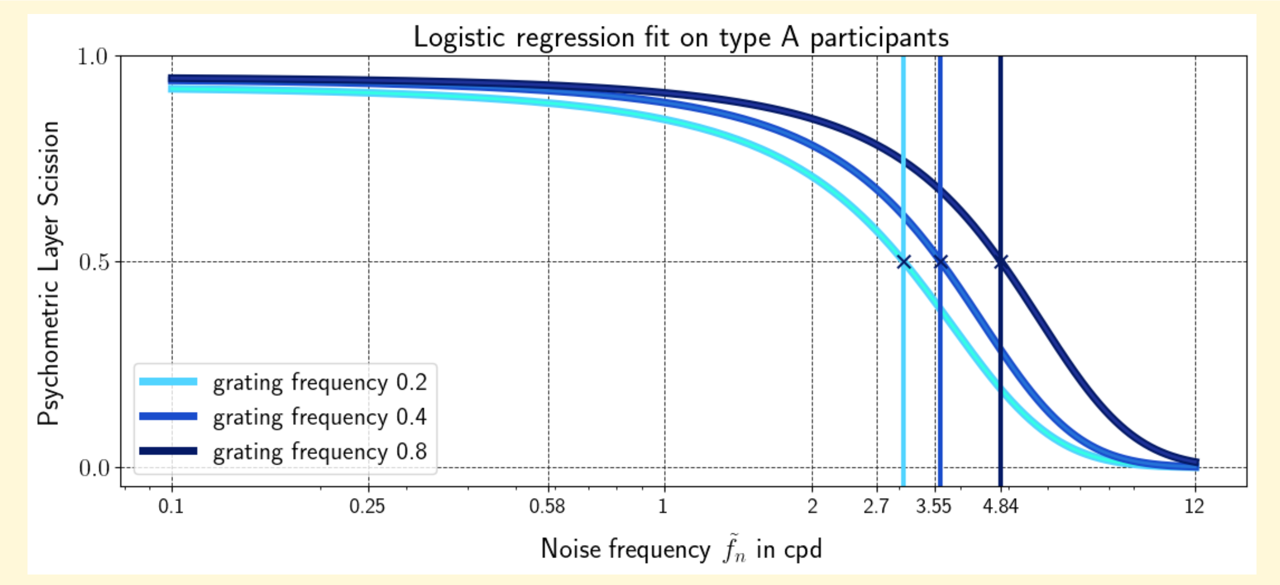


Figure 14: *Fit of logistic regression model on data points of type A observers. One the x-axis noise mask frequency in cpd is shown. y-axis shows the probability of the perception of a layer scission. Varying grating frequencies are again depicted in shades of blue (darker means higher grating frequency). Blue vertical lines indicate critical noise frequencies. The critical noise frequency for* $\tilde{f}_g = 0.2$ *is around* 3.07*, for* $\tilde{f}_g = 0.2$ *is around* 3.65 *and for* $\tilde{f}_g = 0.2$ *is around* 4.84 *The*

We will focus on type A for further analysis since they were the most common and the author falls into this category. We will discuss the difference between observers later in the discussion, but for now, we will investigate which grating frequencies shift at mid frequencies the critical points. We will combine data points of type A observers and then fit a psychometric function with logistic regression.

Psychometric Layer scission

psychometric function A psychometric function gives us the probability of yes based on some variable, such as noise frequency. We will derive three psychometric functions because we have $|f_g| = 3$ grating frequencies. For each psychometric function, the y-axis denotes the yes proportion and the x-axis denotes the noise frequencies. The psychometric function gives us the probability of yes based on noise frequency. We can then see the effect of noise frequency and grating frequency in the form of three psychometric functions. To derive the psychometric layer scission function $LS_{psych\ f_g}$, we fit a logistic function (i.e., a function shaped like an S curve) based on absolute layer scission for type A observers. We did this because the average layer scission is roughly shaped like a S-shaped curve. Observers did binary classification for each image. To estimate the tipping point at which an observer switches between observing and not observing a layer scission for each grating frequency, a simple supervised regression model can be applied to the data. Where the input $\tilde{f}_n$ is a real valued positive number representing the noise frequency in cpd and the label $y \in \{0, 1\}$ describes whether a layer scission is perceived (1) or not (0).

For each grating frequency, we applied logistic regression. We define the psychometric layer scission corresponding to as

$$\begin{aligned} &LS_{psych\ f_g} : \mathbb{R}^+ \to [0, 1], \\ &LS_{psych\ f_g}(\tilde{f}_n) := \frac{1}{1 + e^{-(w_1 \cdot \tilde{f}_n w_2)}} = \text{probability}. \end{aligned} \tag{12}$$

The w_1 and w_2 are learned parameters, where w_1 is the learned coefficient and w_2 is the learned intercept. The function used for this regression task is `LogisticRegression` from `sklearn.linear_model` library. Due to a small data set, the solver `liblinear` has been used and all other parameters were set to default. The data set used to learn w_1 and w_2 contained all data points of type A observers. We used data from participants A_1, A_2, A_3, A_4 and A_5. This meant that there were 100 data points for each noise frequency. The result is plotted in Figure 14

Psychometric layer scission $LS_{psych\ f_g}$ *gives us an estimate of the probability of a given participant voting yes based on noise frequency. This means that if we invert* $LS_{psych\ f_g}$*, we can predict the noise frequency based on probability.*

Estimating critical points To calculate the critical point, the inverse of our prediction function is calculated and the probability y is set to 0.5. For the grating frequency, $0.2cpd$, noise frequency $3.07cpd$ is predicted as the critical point, for grating frequency, $0.4cpd$, noise frequency $3.64cpd$ is predicted and and for $0.8cpd$ noise frequency $4.84cpd$ is predicted. Or in other words

$$LS_{psych\ 0.2}(0.5)^{-1} \approx 3.07, \quad (13)$$
$$LS_{psych\ 0.4}(0.5)^{-1} \approx 3.65 \text{ and} \quad (14)$$
$$LS_{psych\ 0.8}(0.5)^{-1} \approx 4.84. \quad (15)$$

A plot of the three psychometric layer scission functions for each type A observer is in the appendix Figure 14 along with the inverse calculation.

Discussion

We will first interpret our results and then relate how our findings might be relevant to other works.

Interpretation

Interpretation of the effect of the noise frequency At *low* to medium noise frequencies, both layers are visible, and soft X-junctions that preserve contrast polarity can be seen at the edges of White's illusion. Both of these cues hint at the transparency of the noise mask.

At *high* noise frequencies, both layers are still visible, but they merge to form a grainy textured version of White's illusion (see Figure 19 for an example of a textured image). Although junctions still preserve contrast polarity in the stimulus, the size of those junctions is very small at high frequencies (see Figure 15 for a simplified example that shows what is meant by size). It could be that the X-junctions might be too small to be detected as such junctions by our visual system to indicate a layer scission. Because no other cues exist in the image, such as a decrease in contrast at the overlapping region, both layers appear as one.

As the noise frequency increases, the junction's size decreases, and there could be a threshold size around medium frequencies. At *medium* frequencies (around 3cpd), both layers are still visible, but it is ambiguous whether transparency exists or whether they are one layer. Some parts of the noise mask sometimes blend into White's illusion, and edges with higher edge contrast (i.e., edges between dark and light bars) seem likelier to not fuse with the mask. It was difficult to assign the stimulus to either of the two classes because it looked ambiguous. Perception of X-junctions at low noise frequencies, lack of perception of X-junctions at high frequencies, and ambiguity around medium frequencies might hint at a threshold noise frequency that determines the threshold size for X-junctions to be detected by our visual system.

Interpretation of the effect of the grating frequency At *high* and *low* noise frequencies, the grating frequencies did not affect transparency perception across observers (see Figure 13).

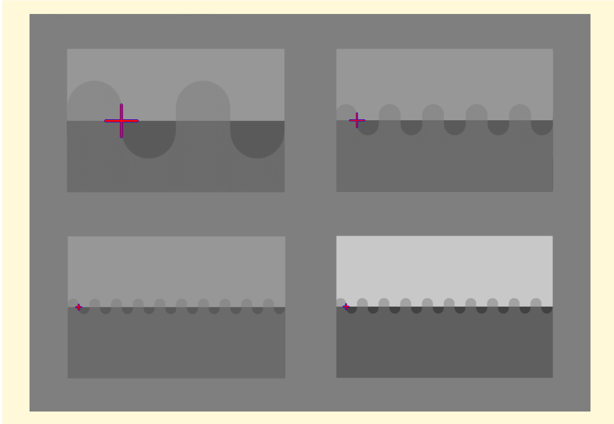


Figure 15: *implified example to show what is meant by changing the size of the +-junction. Top left: A big X-junction can be seen. Top right and bottom left: As the frequency of the pattern of the transparent layer increases the junction size gets smaller. Bottom right: The contrast of the background is increased and the junction size is the same as the bottom left. You might want to move away from the monitor to find a distance at which the structure is perceived at the bottom right but merges at the bottom left.*

However, at *medium* noise frequencies (2 to 5), the threshold for detecting transparency shifted with grating frequency.

While the image contrast stayed the same across different grating frequencies, our sensitivity to it has not stayed the same. The frequency content in images can influence our sensitivity to spatial structure (Bex et al., 2009; Campbell and Robson, 1968). (Campbell & Robson, 1968) showed that the frequency of a squarewave grating can change our sensitivity to spatial structures at the same Michelson contrast. In this experiment, we used White's illusion, which mainly consisted of a squarewave grating. Campbell and Robson (1968) measured (Michelson) contrast sensitivity for squarewave gratings and showed

$$\text{V sensitivity}(0.2) < \text{V sensitivity}(0.4) < \text{V sensitivity}(0.8), \quad (16)$$

where V is the Michelson contrast. This might explain the shift in critical point or threshold noise frequency in our experiment because we derived the following:

$$LS_{psych\ 0.2}(0.5)^{-1} < LS_{psych\ 0.4}(0.5)^{-1} < LS_{psych\ 0.8}(0.5)^{-1} \quad (17)$$

for most observers in their psychometric function (see Figure 14), Figure 21). When contrast sensitivity is higher, we are more likely to see spatial structures. This could lead to small X-junctions near the threshold to more likely to be perceived as such. Therefore, we think that a change in grating frequency changed contrast sensitivity at a fixed noise frequency, making small spatial structures more visible.

It is, however, important to note that this might only explain transparency perception for these stimuli, since X-junctions are not necessary for transparency to be perceived, when other cues are also visible.

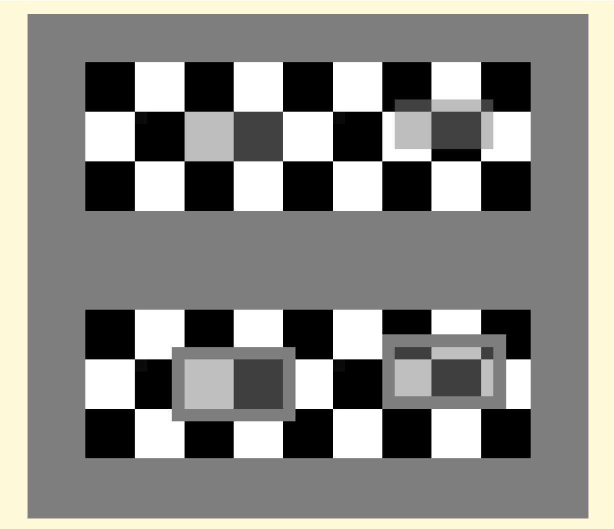


Figure 16: *The top checkerboard is a recreation from Anderson, Barton L. and Singh, Manish (2002): Two rectangular regions have a lower contrast than the rest of the checkerboard. The right rectangle is seen as a transparent sheet on top of the checkerboard, and X-junctions are present. The right rectangle aligns perfectly with the checkerboard but is not seen as transparent. The bottom checkerboard shows the same image but with a grey border around the rectangles to prevent explicit X-junctions. However, the rectangle at the bottom right still looks transparent.*

Figure 16 shows that transparency perception does not always require X-junctions. It can also appear from high-level cues derived from context. In this case, it is reasoned that transparency is perceived because of a decrease in contrast and continuation of checkerboard in the region of overlap (Anderson, Barton L. and Singh, Manish, 2002).

Limitations

The following section lists limitations that might have influenced the results: Sample size and model robustness, and simple labeling give too much freedom to participants, leading to strong deviation between participants.

Not enough grating frequencies We did not test enough grating frequencies to show the direct connection between transparency perception and contrast sensitivity. Although our data shows that the grating frequency induces a shift in critical points similar to contrast sensitivity measured on squarewave gratings in Campbell and Robson (1968), we only used three grating frequencies. Therefore, it is hard to tell how much the threshold relates to contrast sensitivity and transparency perception.

The prediction of exact tipping points The prediction of exact tipping points is not robust enough. With a small data set and high variance within observers, it is almost impossible to predict exact threshold points. Our data indicates that the critical noise frequency appears to be in the interval [2.5,5], yet we only tested two noise frequencies (2.7 cpd and 3.55 cpd) within that interval. Logistic regression relies heavily on the sample size of the data and other hyperparameters. For example, choosing a function to fit the data and choosing which error minimization. These conditions influence the outcome of the prediction.

Experiment method Another problem might also be the methodology used to collect our data set. Most participants followed the expected pattern: yes, transparency at low noise frequencies, maybe transparency at medium noise frequencies, and no transparency at high noise frequencies. Some observers deviated from this pattern. Two participants saw transparency in every condition. They explicitly explained that they saw clear and sharp edges at the grating, which made them see transparency. This could be due to a higher acuity in their vision and therefore can better detect small spatial strucures. However, this case is highly unlikely since all of the experienced observers saw no transparency at high frequencies, mentioning that they saw fuzzy edges. This discrepancy might be due to the type of task we chose since the yes-no format might have implied to some participants that they were being tested on how good they were at detecting transparency. Additionally, since many conditions contained perceptual transparency, they might have seen transparency due to that. This might be minimized by masking the entire screen instead of embedding a masked stimulus on a gray background. And also adding a control condition, in which no one should perceive transparency. For example, occasionally displaying White's illusion without the mask at a lowered contrast. Additionally, one might use a more nuanced approach, such as showing two images of different parameters and asking which image evokes a stronger impression of transparency or which image seems more likely to have transparency. Or adjust parameters dynamically until they do not see a layer scission, like dynamically increasing noise frequency until a no is pressed. Some participants might not have understood the task or had other (not expected) contextual ideas. AB initially had an expected response but, after a break, suddenly saw transparency in every condition. He stated that the order of layers appeared reversed (White's illusion in front of the noise mask). C1 initially correctly explained transparency before the experiment and seemed to understand the task, but afterward, she could not explain her reasoning. One of the causes might be that she mistook left for right because before the experiment started, the text displayed on the lab monitor incorrectly switched left and right.

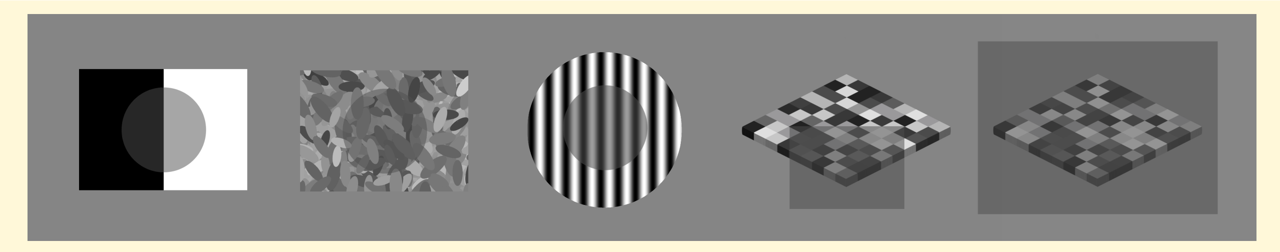


Figure 17: *Recreation of stimuli from left to right: All images except e) contain a filter of the same reflectance and transmittance. Obvious cues like x-junctions are visible in all images except the last. a) A round disk (epicoster rotating at high enough speed such that it looks transparent) in front of two aligned rectangles.(Metelli, 1974); b) A small moving filter across the background in front of a dissimilar background consisting of hundreds of ellipses varying in luminance, rotation, and length. Robilotto and Zaidi, 2004; c) A small filter in front of a sinewave grating. Anderson, Barton L. and Singh, Manish, 2002; d) A rectangular filter in front of a variegated checkerboard. Geometric cues hint that the filter is not coplanar to the checkered slab. Aguilar and Maertens, 2022; e) Modified version of Aguilar and Maertens (2022), now the filter is larger, covering the entire slab. Without context, this image appears just darker and with less contrast. Because of the layout in this illustration and context (i.e., hint that in every image, there is a gray filter in front of something), one might still see a filter in front of the image.*

Related

Comparison to Betz et al. (2015a) and limitation of comparison One of the side goals of this thesis came from the origin of the stimuli from Betz et al. (2015a). Because participants in Betz et al. (2015a) informally observed transparency shift based on noise frequency, we wanted to compare how the critical points of perception of transparency relate to the critical points for illusion strength from Betz et al. (2015a). In Betz et al. (2015a), the task of the participants was to match the lightness of the target patch, which was located next to the illusion, to the test patch within the White's illusion. For each match, they measured the illusion strength, which describes how much the actual luminance of the test patch in the illusion before masking deviates from the matched luminance of the participant (i.e., big deviance in luminance, then the illusion strength big). We can only roughly compare those points because of the mentioned Limitations. Roughly, the points are in the same region, here 3 to 5 cpd and in Betz et al. (2015a) 2 to 4 cpd. Also, both points are not equidistant spaced. It could be that both are following the same underlying mechanism shifted by some other mechanism or variable. The underlying mechanism could be edge detection. X-junctions require edges (can be soft), and in White's illusion, it has been shown that edges play an important role in perceiving lightness in this illusion. However, not much more can be interpreted.

Limitation of comparison Additionally replicating the stimuli with the same contrast was difficult. It was difficult to replicate the same stimuli without increasing the Michelson contrast of the grating, and the grating was barely visible for some noise conditions. Notably, Betz et al. (2015a) had similar problems when recreating stimuli from Salmela and Laurinen (2009). Betz et al. (2015a) mention that in their early replication, observers did not see the chest patches and therefore made it impossible to judge the lightness on the patch. Because of that they changed the contrast. On the one hand, it further underlines the high sensitivity of our visual system at 2 to5 cpd for spatial structures. On the other hand, it is a problem when recreating experiments that contain noise masks because it also makes it difficult to replicate stimuli with noise masks since small changes, like contrast, might lead to notable differences.

Relation to transmittance In this thesis, we investigated when transparency is perceived. However, a lot of researchers also investigated how *transparent* a layer is perceived (i.e., how *transmissive* one layer looks) and there have been many attempts to predict perceived transmittance (Metelli, 1974; Robilotto and Zaidi, 2004; Anderson, Barton L. and Singh, Manish, 2002; Aguilar and Maertens, 2022). Similar to Metelli (1974) we could model the stimuli from our experiment that were created by

$$\texttt{stim["img"]:=} \quad \begin{array}{c} \texttt{0.5 * noise["img"]} \\ + \\ \texttt{0.5 * whites["img"]} \end{array} \tag{18}$$

with Talbot's law. This law states

$$P = \alpha a I + (1 - \alpha) t I,$$

where P is the resulting Luminance, α transmittance of transparent layer, a reflectance of background layer and t reflectance of transparent layer. For each point in the image as

$$s_{i,j} = 0.5 a_{i,j} + (1 - 0.5) t_{i,j},$$

where $s_{i,j}$ is luminance of stimulus, $a_{i,j}$ luminance of White's illusion and $t_{i,j}$ luminance of transparent mask, and i, j denotes the position. Under this model, the mask has a transmittance $\alpha = 0.5$.

Implications and Suggestions for further research

Further investigate how contrast sensitivity affects transparency Our experiment might imply that changing the frequency of dissimilar backgrounds changes the perceived transmittance of the transparent medium. Perceived transmittance seems to closely relate to perceived contrast (Robilotto and Zaidi, 2004; Aguilar and Maertens, 2022). Stimuli used to investigate perceived transmittance contained obvious cues like X-junction, and a decrease in contrast in the region of overlap can be seen (see Figure 17). Notably, these stimuli contain patterned backgrounds. However, none of these studies changed the frequency of the background pattern. Changing the spatial frequency content changes contrast sensitivity. This implies that changing frequency content might change the perceived transmittance. To test this, one might use modified versions of existing stimuli. For example, modified stimuli from Anderson, Barton L. and Singh, Manish (2002) could be used since a contrast sensitivity function from Campbell and Robson (1968) exists for the sinewave grating.

Finding a second tipping point at low noise frequencies Our experiment found one tipping point for each grating frequency at noise frequencies between 3.1 to 5 cpd. As noise frequency approaches 0 *cpd*, changes in luminance become more gradual until the noise mask appears as one even surface. Therefore, there is most likely a second tipping point and noise frequencies lower than 0.1 cpd. Where is the tipping point, and does the tipping point also shift with grating frequency?

Using a different types of noise masks In our experiment, we only used narrowband noise. However, it has been shown that different types of noises can interfere with edges differently Sørensen, 2023. Thus, spatial structures are perceived differently. Therefore, one could investigate how different the critical points are between different types of noises.

Conclusion

In this study, we adjusted White's illusion from Betz et al. (2015a) and investigated at which point the noise mask causes a layer scission. The way the mask was merged with the illusion in both experiments made it possible to perceive a layer scission in the form of transparency. To investigate under which conditions transparency is perceived, we conducted an experiment that tested varying noise frequency, grating frequency, and grating phase. We found out that at low noise frequencies, transparency is perceived, and at high frequencies, transparency is not perceived in most participants and that the switch from yes to no appeared in noise frequency between 2.5 to 5. As the grating frequency increased, the critical noise frequencies increased. This increase, similar to Betz et al., 2015a, is not proportional. We argue that soft X-junctions cause the noise mask to initiate a layer scission in the form of transparency at low frequencies (0.1 *cpd* to 2 *cpd*). We think that because the noise mask covers the entire White's illusion and part of the background, we argue that X-junctions and consistency of both layers are the only cues in this experiment for transparency. When this cue X-junctions is removed, the layer scission is gone. Additionally, the grating frequency shifted these points, and we conclude that this might be due to a change in contrast sensitivity induced by a change in grating frequency. This undermines the fact that frequency content might be important for low-level cues for transparency, even though it has been left out of many studies investigating transparency perception. Studies investigating how transmissive a transparent layer appears often use obvious cues like X-junctions to initiate transparency perception and then investigate how transmissive the transparent layer looks to derive a model and metric to predict human behavior. Even though it could be a critical variable in modeling, it remains an open question to what extent frequency content and contrast sensitivity relate to that.

Acknowledgements

I want to thank Dr. Guillermo Aguilar for first introducing me to the topic of visual perception when I visited the seminar *Visuelle Wahrnehmung* and for helping me with setting up the code for the experiment. Prof. Dr. Marianne Maertens for giving useful advice before writing this thesis. Bianca del Mestre for managing administrative activities such as giving me access to the student working room and adding the work of bachelor students including me to the psyco tu Berlin website and the group at Computational Psychology in general for creating a fun environment such as planning a student symposium.

Use of Ai

Because of Eigenständigkeitserklärung, I will list which AI tools I used and how I used them. I used ChatGPT for subtasks programming in Python and writing LaTeX code because it was faster than googling, and the interface looked cleaner than most websites. This did not replace reading documentation. Some prompts I used were:
how do i add a sticky note type box on a plot?
or
how do i \printbibliography in alphabetically order?
These often resulted in code snippets that had been

heavily changed since they usually worked but not how I intended. In the early stages of writing, I used ChatGPT (Version 4o) to check the correctness of grammar; however, instead of correcting the grammar, it often twisted my words and made incorrect statements, so I switched to Grammarly. I used that to check the correctness of the grammar of my written text (punctuation, spelling, capitalization). Occasionally, I used the speech-to-text feature of ChatGPT (Version 4o) to reduce eye strain and better express my thoughts. None of the images were AI-generated; they were either generated in Python, drawn digitally in Procreate on my iPad, or photographed.

References

Aguilar, G., & Maertens, M. (2022). Conjoint measurement of perceived transparency and perceived contrast in variegated checkerboards. *Journal of Vision*, *22*(2), 2. https://doi.org/10.1167/jov.22.2.2

Anderson, B. L. (1997). A Theory of Illusory Lightness and Transparency in Monocular and Binocular Images: The Role of Contour Junctions. *Perception*, *26*(4), 419–453. https://doi.org/10.1068/p260419

Anderson, Barton L. & Singh, Manish. (2002). Toward a Perceptual Theory of Transparency.

Betz, T. (2016). Arguments for a new early vision model of lightness perception. Retrieved April 20, 2024, from https://depositonce.tu-berlin.de/handle/11303/5429

Betz, T., Shapley, R., Wichmann, F. A., & Maertens, M. (2015a). Noise masking of White's illusion exposes the weakness of current spatial filtering models of lightness perception. *Journal of Vision*, *15*(14), 1. https://doi.org/10.1167/15.14.1

Betz, T., Shapley, R., Wichmann, F. A., & Maertens, M. (2015b). Testing the role of luminance edges in White's illusion with contour adaptation. *Journal of Vision*, *15*(11), 14. https://doi.org/10.1167/15.11.14

Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision*, *9*(10), 1. https://doi.org/10.1167/9.10.1

Campbell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, *197*(3), 551–566. Retrieved December 4, 2024, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1351748/

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., & Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*(4), 795–834. https://doi.org/10.1037/0033-295x.106.4.795

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy [Publisher: Nature Publishing Group]. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Howe, P. (2005). White's Effect: Removing the Junctions but Preserving the Strength of the Illusion. *Perception*, *34*, 557–64. https://doi.org/10.1167/5.8.562

Maertens, E. D. M., & Wichmann, Z. D. F. (n.d.). Ist das visuelle System des Menschen empfindlicher für Kanten oder Luminanzunterschiede?

Metelli, F. (1974). The Perception of Transparency. https://doi.org/10.1038/scientificamerican0474-90

Michelson, A. A. (1995, January). *Studies in Optics* [Google-Books-ID: m2tUZ4_8WGMC]. Courier Corporation.

Robilotto, R., & Zaidi, Q. (2004). Perceived transparency of neutral density filters across dissimilar backgrounds. *Journal of Vision*, *4*(3), 5. https://doi.org/10.1167/4.3.5

Salmela, V. R., & Laurinen, P. I. (2009). Low-level features determine brightness in Whites and Benarys illusions. *Vision Research*, *49*(7), 682–690. https://doi.org/10.1016/j.visres.2009.01.006

Schmittwilken, L., Maertens, M., & Vincent, J. (2023). Stimupy: A Python package for creating stimuli invision science. *Journal of Open Source Software*, *8*(86), 5321. https://doi.org/10.21105/joss.05321

Sørensen, J.-S. (2023). The Inuence of Noise on Human Edge Perception in Natural Images.

White, M. (1979). A New Effect of Pattern on Perceived Lightness. *Perception*, *8*(4), 413–416. https://doi.org/10.1068/p080413

Zaidi, Q., Spehar, B., & Shy, M. (1997). Induced effects of backgrounds and foregrounds in three-dimensional configurations: The role of T-junctions. *Perception*, *26*(4), 395–408. https://doi.org/10.1068/p260395

Appendix

Calculations and proofs Calculating the inverse function can be archieved by solving for x.

$$
\begin{aligned}
&& y &= \frac{1}{1+e^{-(ax+b)}} \\
\Leftrightarrow && y(1+e^{-(ax+b)}) &= 1 \\
\Leftrightarrow && ye^{-(ax+b)} &= 1-y \\
\Leftrightarrow && \ln(y)-(ax+b) &= \ln(1-y) \\
\Leftrightarrow && -ax-b &= \ln(1-y)-\ln(y) \\
\Leftrightarrow && -ax-b &= \ln(\frac{1-y}{y}) \\
\Leftrightarrow && x &= \frac{1}{a}(\ln(\frac{1-y}{y})+b)
\end{aligned}
$$

learned parameters of the model are: Parameters for critical point for 0.2 is ≈ 3.07 and parameters are

$$w_1 \approx -0.82186902 \quad \text{and} \quad w_2 \approx 2.5233571$$

Parameters for critical point for 0.4 is $\approx$ 3.65 and parameters are

$$w_1 \approx -0.77676069 \quad \text{and} \quad w_2 \approx 2.83494418$$

Parameters for critical point for 0.8 is $\approx$ 4.84 and parameters are

$$w_1 \approx -0.60318728 \quad \text{and} \quad w_2 \approx 2.91922782$$

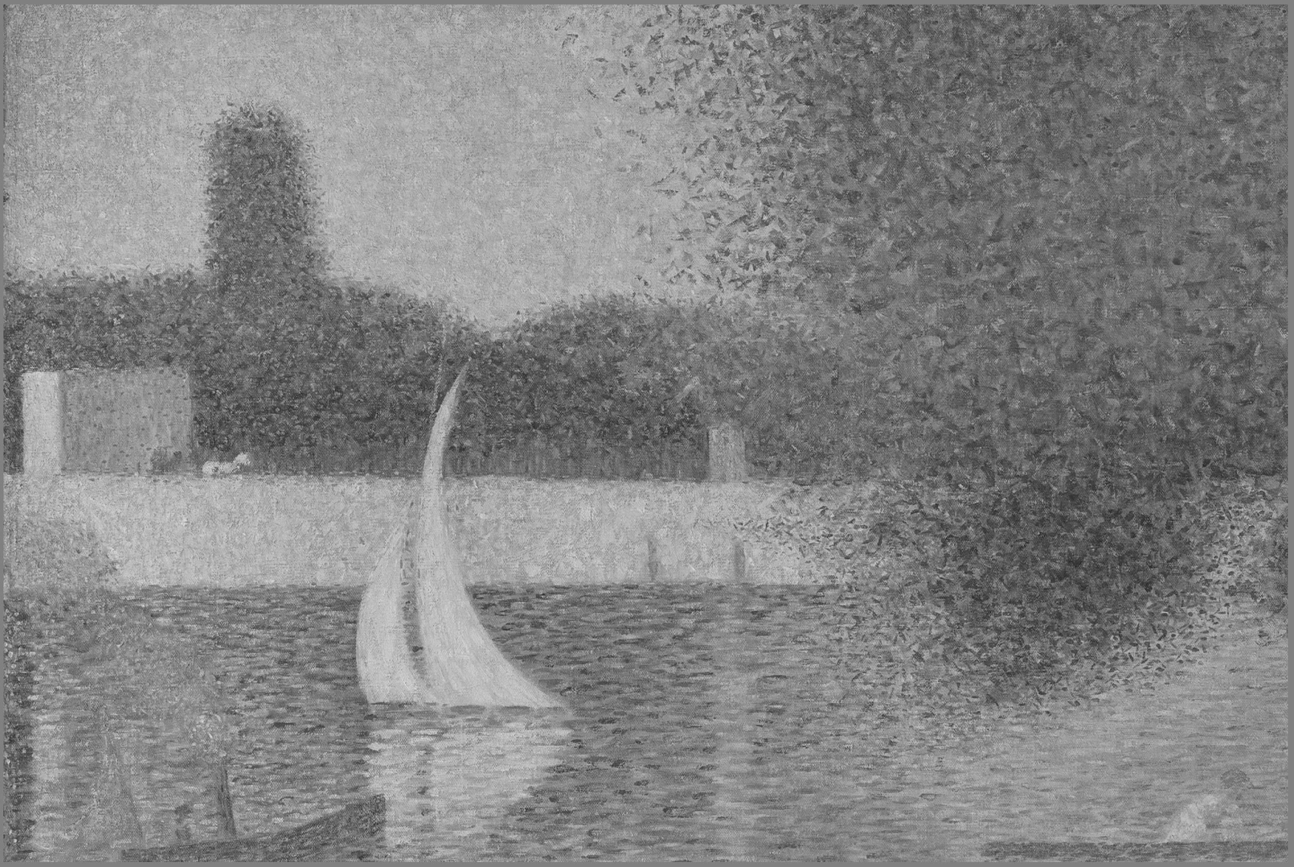


Figure 18: *Georges Seurat, A sunday on La Grande Jatte 1884, enlarged and edited. Personal comment: Pointilism is perceived as one layer and looks similar to stimuli with noise at 9cpd.*

Figure 19: *A painting composed of thousands of sand corns.*

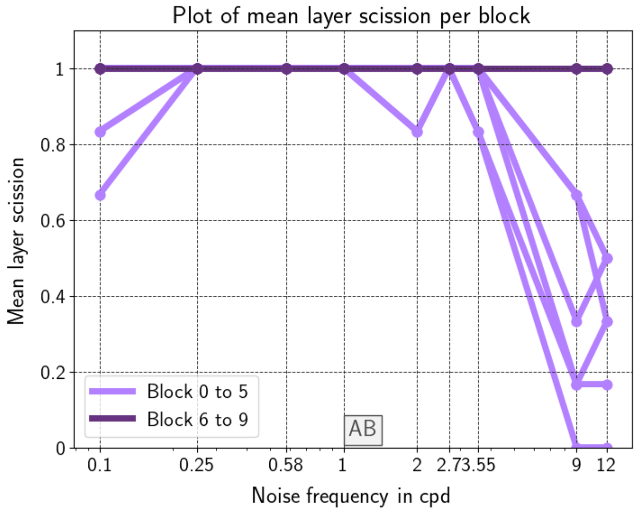


Figure 20: *On the x-axis is noise frequency in cpd and on the y-axis is the mean of layerscission grouped by noise frequency and block. Light shades indicate that the block is in* $\{0,1,2,3,4,5\}$ *and dark shades in* $\{6,7,8,9\}$*. For simplification for this visualization there is no grouping by the three grating frequencies. Each dot shows the mean of* $|\varphi_g| \times |f_g| = 2 \times 3 = 6$ *button presses in one block.*

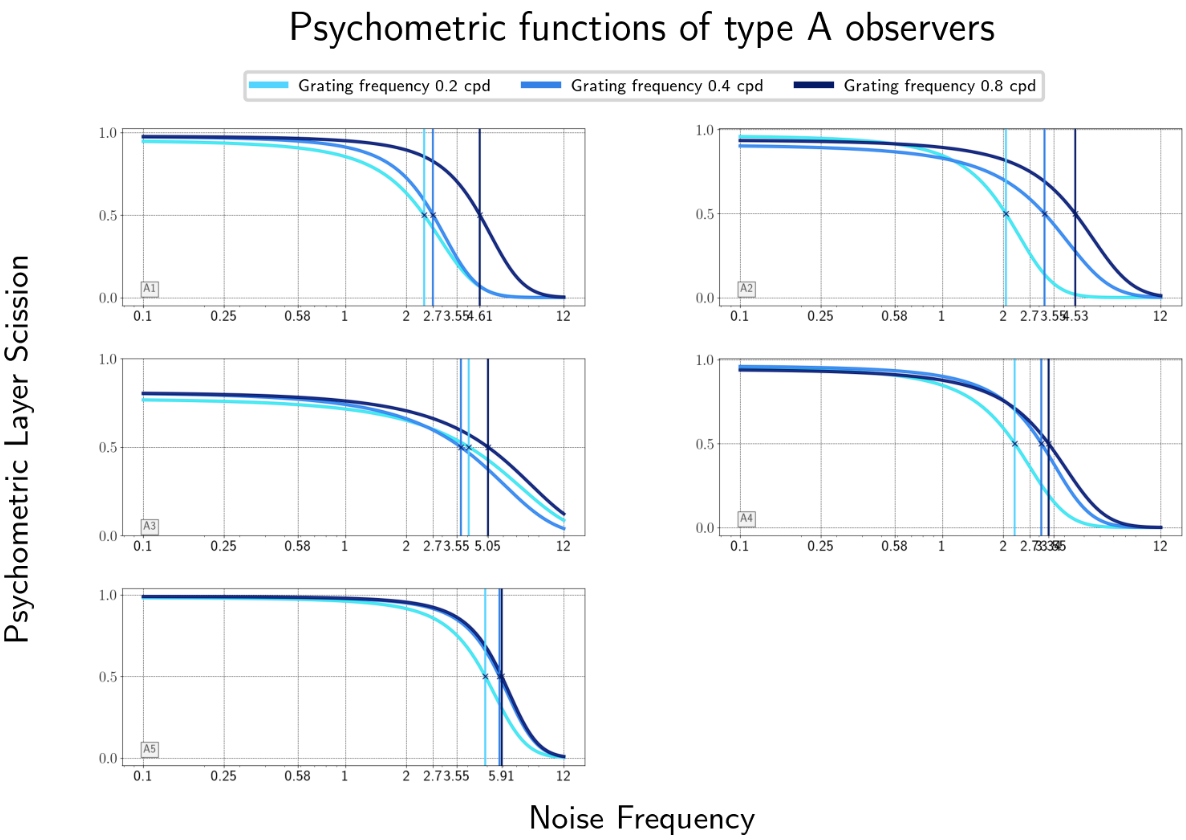


Figure 21: *Psychometric layer scission for each type A observer. Darker blue shades indicate higher grating frequency (0.2,0.4,0.8). Vertical lines indiacate critical points. Grey sticky notes indicate participant alias. Notably, only half of the participants had the critical point shifted in ascending order of grating frequency.*

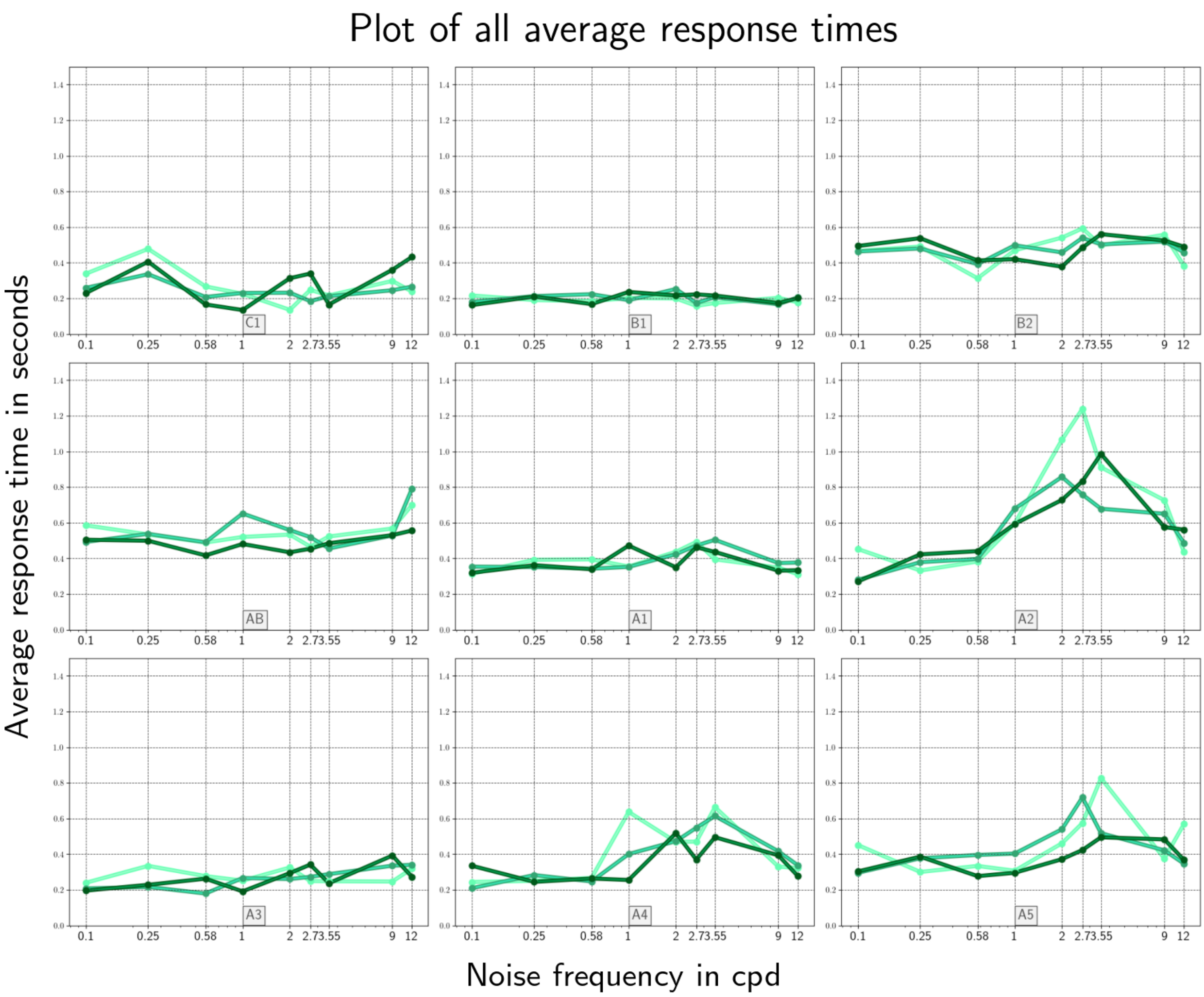


Figure 22: *Average response time from all participants. Darker green shade indicates higher grating frequency (0.2,0.4,0.8). Intentionally changed to green to distinguis that this data was not used for layer scission analysis. Grey sticky notes indicate participant alias.*

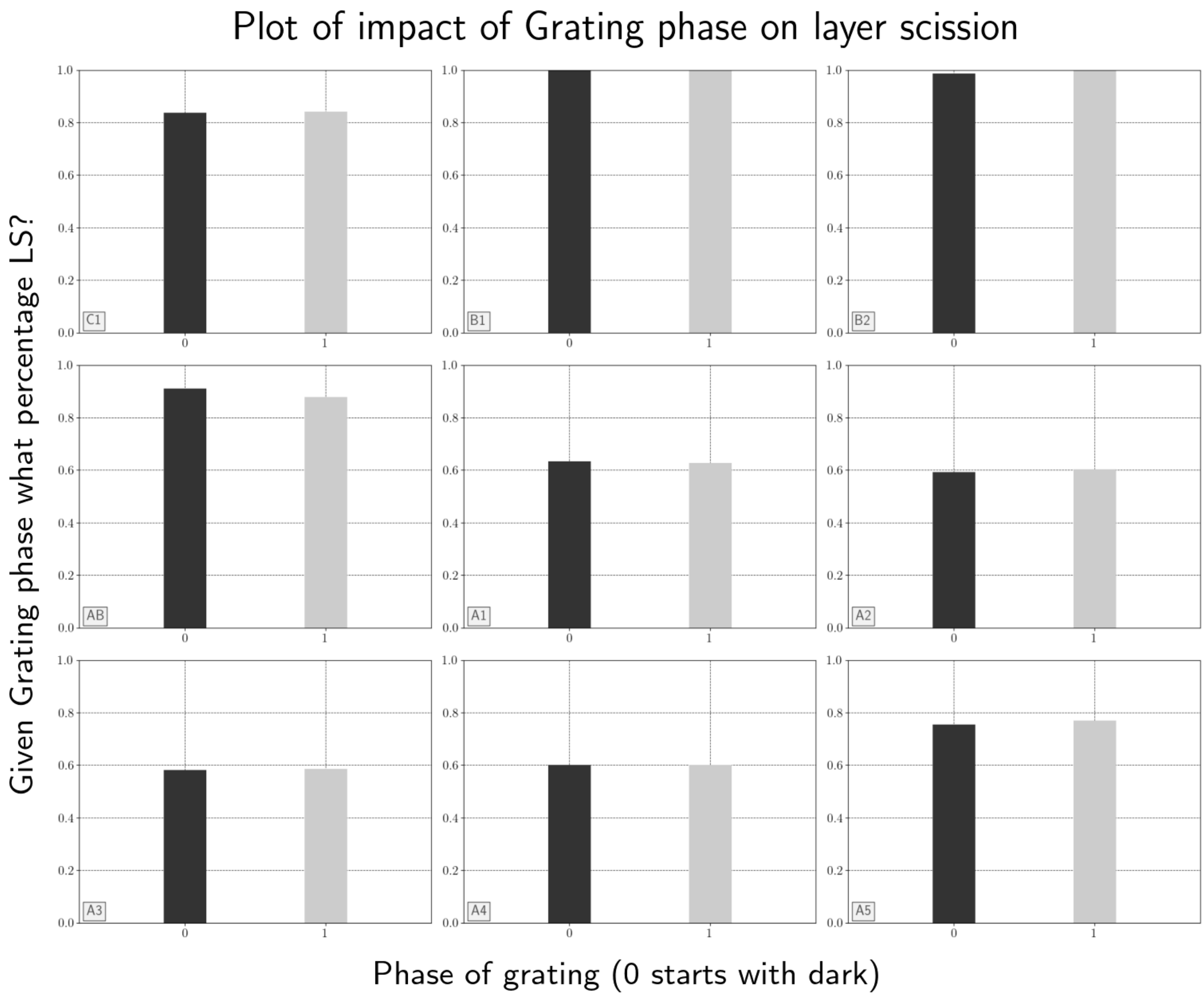


Figure 23: *Average impact of grating phase on layer scission. Dark bars indicate that White's illusion started with a dark bar and light bars indicate that it started with a light bar.*

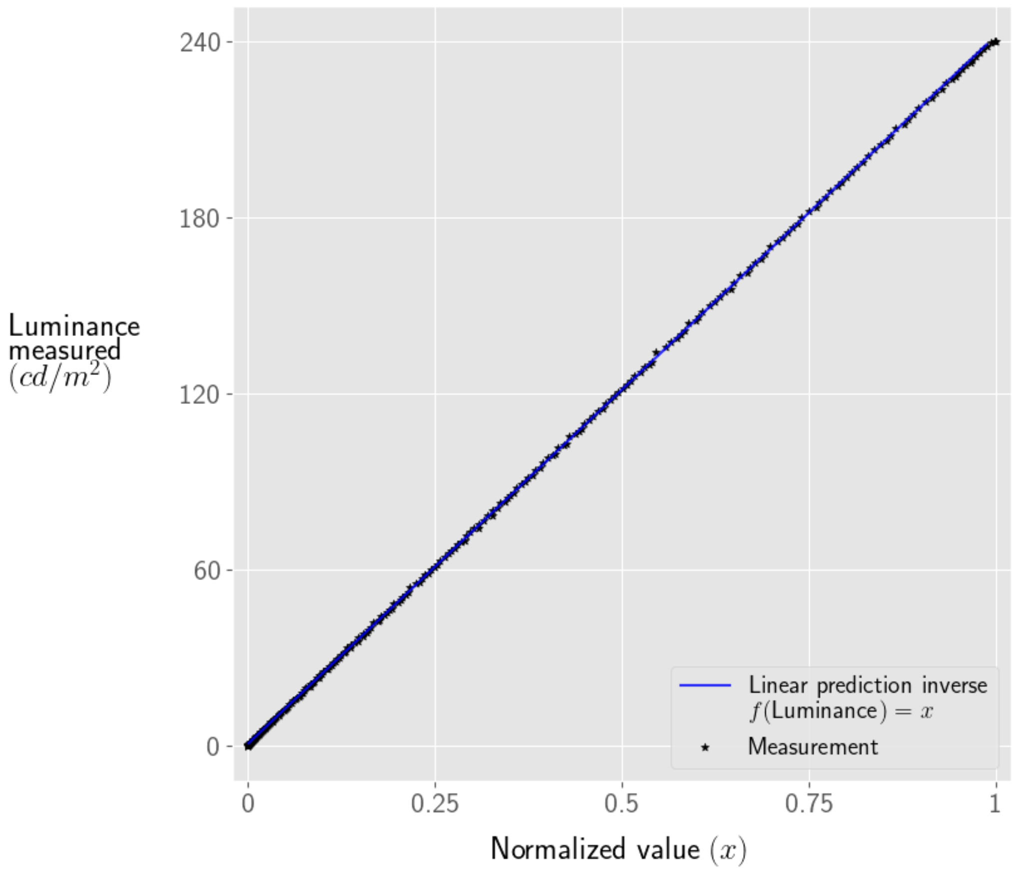


Figure 24: *The x-axis shows the displayed normalized values from* 0 *to* 1 *and y-axis shows the measured luminance in* cd/m^2. *The black stars are the individual measurements and the blue line depicts the predicted normalized value based on luminance.*

Experiment Instruction

Hannah Louisa Boldt

What is a layer scission?

Generally a layer scission is a parting of an image into multiple overlapping *layers*. For example, clouds in front of a mountain (see Figure 4).

Figure 1: Simplified example. Left side: no layer scission is observed since the squares sit next to each other in the same plane. Right side: Layer scission can be observed since the light grey square appears *transparent* (Meaning it lets light pass through) and in front of the other square

Please alsolook at the follwowing examples on the next page for more context.

Instruction

You will be shown simplified images in a randomized order and have to judge for each image whether you perceive layer scission or not.

(i) Experiment starts

   (i) Image is shown for a **short time** (500 ms) (Please look at the middle of the image, the middle is indicated by a black dot after the image is gone)

   (ii) When the image is gone, the system waits for a response. You can either **press RIGHT (→) to indicate layerscission** or **press LEFT (←) to indicate *no* layerscission**. You cannot undo your judgement.

   (iii) The next image is immediately shown and the sequence is repeated!

(ii) This sequence repeats 10 times with short breaks in between.

The experiment itself will take around 40 minutes.
There is no correct answer when judging the images. Please choose the option that looks closest to what you perceive.
It might happen, that you are very unsure for some, none or a lot of images. This can happen. If you are unsure, choose the answer that looks more sound at that moment.

Example images

Figure 2: No layer scission can be seen because the carpet itself is colored. IKEA Carpet

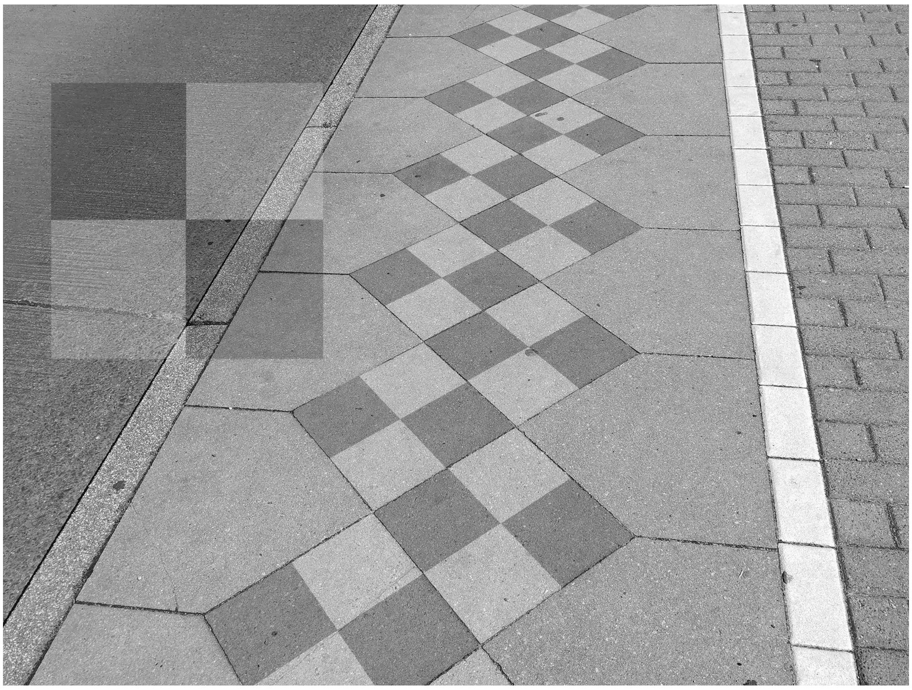


Figure 3: For the left square a layer scission can be observed since the square is transparent and infront of the image. Squares on the right (path) are not seen as a layer scission since the stones are colored.

2

Figure 4: Example of a layer scission (with respect to image content), the clouds are in front of the mountain.

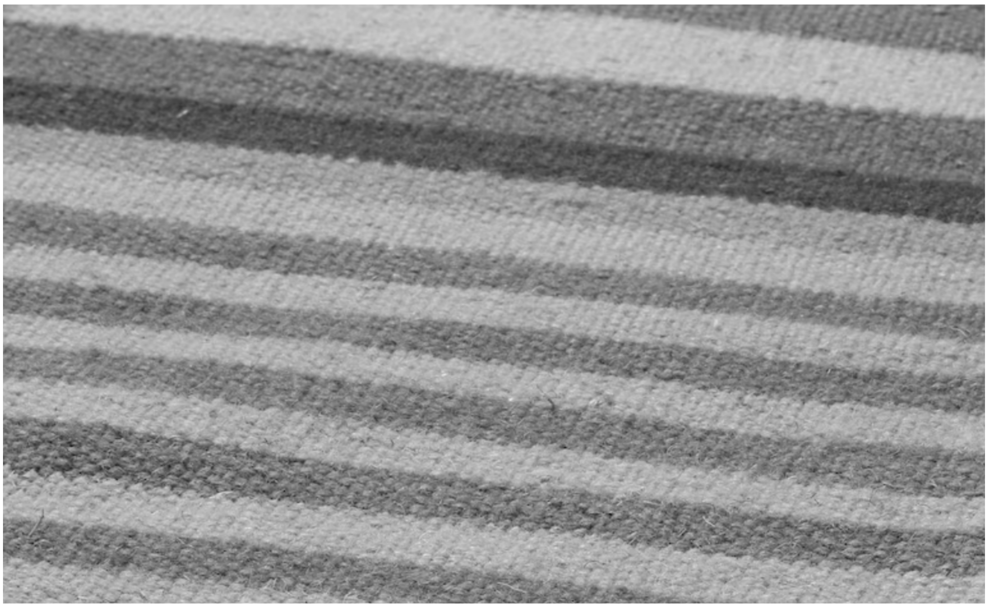


Figure 5: No layer scission can be seen because the carpet itself is colored. IKEA Carpet

3

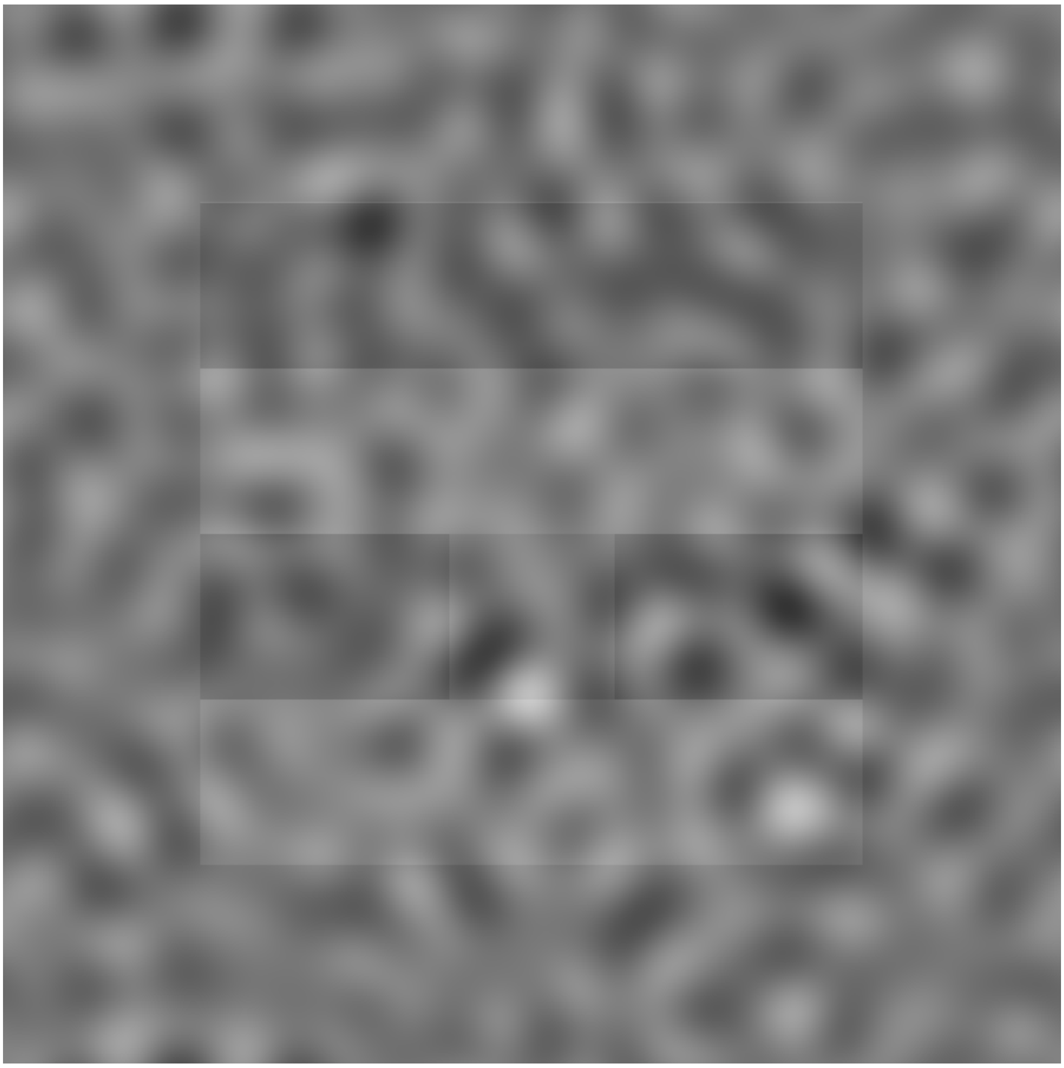


Figure 6: Example *Stimulus* that will be shown in this experiment. ↑

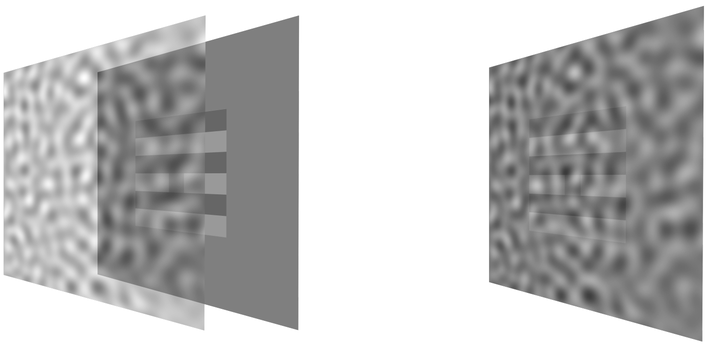


Figure 7: Illustration of perceptual categorization of the stimulus in multiple layers of different transparency. Left: The stimulus is perceived as two layers. The noise is seen as a transparent layer in front of the White's illusion. Right: The stimulus is perceived as one layer. The noise is seen as part of the White's illusion, similar to objects with texture.

4