Technische Universität Berlin
Electrical Engineering and Computer Science
Institute of Computer Engineering and Microelectronics
Computational Psychology

Investigating Different Methods to Study Human Brightness Perception

Hafidz Arifin

Thesis submitted in fulfillment of the requirements for the degree
"Bachelor of Science" (B. Sc.) in the field of Computer Science

November 2023

First examiner: Prof. Dr. Marianne Maertens
Second examiner: Prof. Dr. Felix Wichmann
Supervisor: Dr. Joris Vincent

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Die selbstständige und eigenhändige Ausfertigung versichert an Eides statt

Berlin, den	10. November 2023	
	Datum	Unterschrift

Zusammenfassung

Ziel dieser Arbeit ist es, die menschliche Helligkeitswahrnehmung zu untersuchen, indem die Beziehung zwischen den Ergebnissen von zwei verschiedenen Messmethoden - Methode der Reizfindung und der Helligkeitsbewertung - untersucht wird. Insbesondere wird untersucht, ob die mit diesen Techniken gewonnenen Daten ähnliche Schätzungen der wahrgenommenen Helligkeit bei verschiedenen Helligkeitstäuschungen liefern. Bei der Methode der Reizfindung müssen die Teilnehmer die Leuchtdichte eines Reizes so anpassen, dass sie einer bestimmten Helligkeit entspricht, während bei der Helligkeitsbewertung die Teilnehmer bewerten, welche Seite des Reizes als heller wahrgenommen wird.

In früheren Studien wurde jeweils eine dieser Methoden zur Messung der menschlichen Helligkeitswahrnehmung einzeln angewendet, wobei die Daten Ähnlichkeiten aufwiesen. Da in diesen früheren Studien jedoch Stimuli mit unterschiedlichen Eigenschaften verwendet wurden und verschiedene Teilnehmergruppen eingesetzt wurden, könnte der Vergleich eher durch Stimuluseigenschaften oder die Individualität der Teilnehmer als durch die Methodik allein beeinflusst worden sein.

Die vorliegende Studie zielt darauf ab, diese Einschränkungen zu überwinden. Sie führt ein Experiment durch, bei dem dieselbe Gruppe von Teilnehmern mit derselben Helligkeitstäuschung konfrontiert wird, und misst dann die Wahrnehmung mit den beiden oben genannten Methoden. Die Ergebnisse zeigen, dass bei bekannten Stimuli ein hohes Maß an Ähnlichkeit hinsichtlich der durchschnittlichen Wirkungsrichtung besteht. Bei neuartigen Stimuli ist diese Ähnlichkeit hingegen gering. Bei der Bewertung der Stärke und Variabilität der Effekte über alle Stimuli hinweg besteht eine allgemein geringe Ähnlichkeit. Entscheidend ist, dass bei der Mehrzahl der bekannten Helligkeitstäuschungen die Ergebnisse hinsichtlich der durchschnittlichen Wirkungsrichtung mit früheren Untersuchungen übereinstimmen, was die Ähnlichkeit der Ergebnisse der beiden Messverfahren für diese Stimuli bestätigt.

Abstract

This work aims to investigate human brightness perception by examining the relation of the results from two distinct methods of measurement - the method of adjustment and brightness ratings. Specifically, the study investigates whether data acquired using these techniques provide similar estimates of perceived brightness in various brightness illusions. The method of adjustment requires participants to adjust the luminance of a stimulus to correspond to a particular brightness, whereas brightness ratings involve participants rating which side of the stimulus is perceived as brighter.

Past studies individually adopted each of these methodologies to measure human brightness perception, with their data displaying similarities. However, since these prior studies utilized stimuli with diverse properties and engaged different participant groups, the comparison might be influenced by stimulus characteristics or individual differences rather than methodology alone.

This study aims to overcome such limitations. It conducts an experiment wherein the same subset of participants is presented with the same brightness illusion and then measures perception using both the methods mentioned earlier. The findings reveal that, for known stimuli, there is a high degree of similarity concerning the average direction of effect. Conversely, for novel stimuli, this similarity is low. When evaluating the strength and variability of effects across all stimuli, there exists a low degree of correlation. Crucially, for the brightness illusions that were also used in prior studies, the results regarding average direction align with results from prior research, thus confirming for these stimuli the similarity of results produced by the two measurement techniques.

Acknowledgement

First, I want to express my gratitude to my supervisor Dr. Joris Vincent for his supportive guidance, valuable advice, and patience. I am grateful for every constructive feedback and enlightening encouragement that I got from him which helped me to write this thesis.

I also want to thank Prof. Dr. Marianne Maertens for her illuminating advice and for giving me the precious opportunity to write my thesis in her department in the first place.

I would also like to thank Prof. Dr. Felix Wichmann for his willingness to be my second examiner.

Finally, I want to thank my family, friends, and everyone else, without whom I would never have gotten this far.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Brightness perception refers to the subjective experience of light intensity which is not solely based on the actual physical measurement of light. *Illuminance*, *reflectance*, and *luminance* are physical properties that influence our perception of surfaces (Adelson 2000). As shown in Figure 1.1, a light source, such as the sun, sends an amount of light which is called *illuminance* to a surface. *Reflectance* is the ratio of the amount of light that is reflected by the surface. The last one, *luminance*, refers to the amount of visible light that reaches to the eye from a surface.

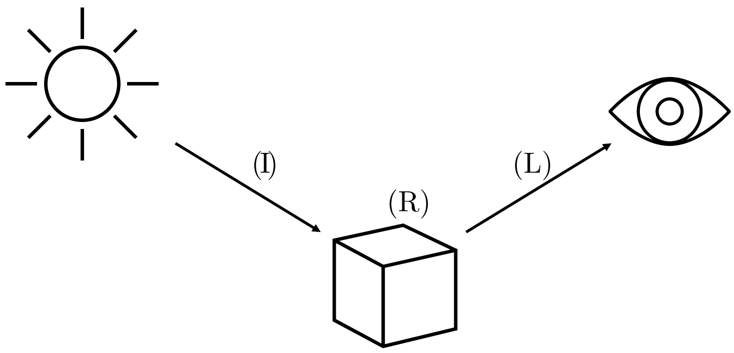


Figure 1.1: The relationship between three physical properties of light and surfaces: Illuminance (I), reflectance (R), and luminance (L). The sun at the top left sends illuminance to an object located at the bottom center. This object reflects its reflectance. The combination of illuminance and reflectance is the luminance, which reaches the eye at the top right. This Figure was inspired by Brunn (2020).

In contrast to these three physical properties, *lightness* and *brightness* are perceptual properties that are subjective. *Lightness* is the perceived reflectance and *brightness* is the perceived luminance. It is essential to distinguish between brightness (a subjective measure) and luminance (an objective measure). An example that shows this difference is the checker shadow illusion by Adelson (1995) shown in Figure 1.2. It shows two tiles labeled as ‘A’ and ‘B’ on a checkerboard. Many would say, the two tiles have two different brightnesses. Upon examining their luminance values, it becomes evident that both tiles have the same luminance. Since Figure 1.2 is a picture, it does not have a reflectance. Assuming the elements in Figure 1.2 are real objects, then tile ‘A’ has a lower reflectance than tile ‘B’ because tile ‘A’ is darker than tile ‘B’. But, on the other hand, tile ‘A’ has a higher illuminance than tile ‘B’ because tile ‘A’ is not covered by a shadow like tile ‘B’. Both tiles have different illuminance and reflectance values, but their mathematical product, which is the luminance, is the same. Tile ‘B’ is located in the shadow and stands out as light due to the presence of darker surrounding tiles that are also partially covered by the shadow. This illusion exemplifies how surrounding context influences the human brightness perception. Hence, the checker shadow illusion demonstrates that human brightness perception is not always congruent with objective luminance, underlining the complex interaction between physical light properties, which are illuminance, reflectance, and luminance, and subjective visual experience, which are lightness and brightness.

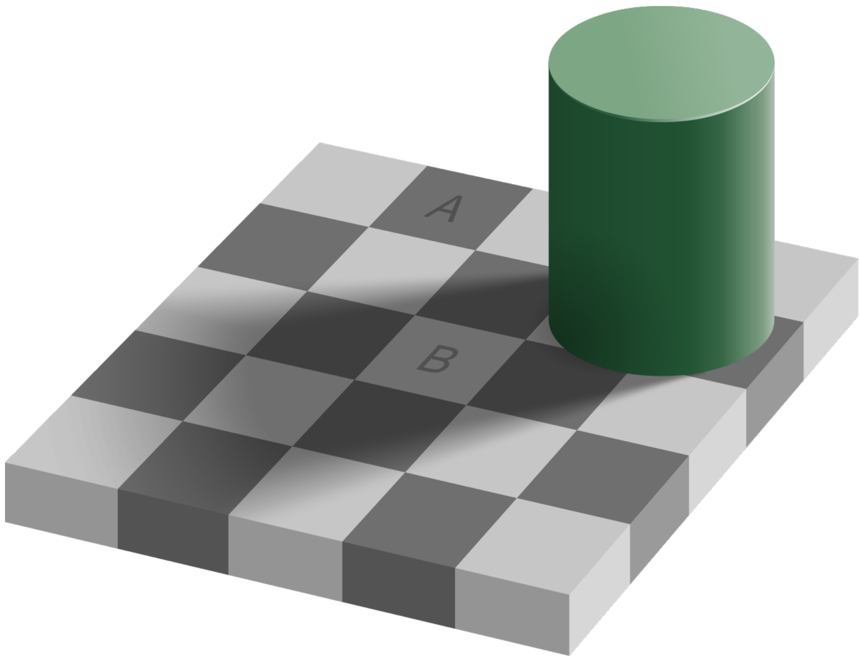


Figure 1.2: Checker shadow illusion. A three-dimensional checkerboard surface consists of bright and dark gray tiles. Two of the tiles are marked with ‘A’ on top and ‘B’ in the middle. On the right side, there is a green cylinder standing on the checkerboard, and it cast a shadow on the checkerboard’s center. The shadow covers and surrounds the tile ‘B’. Figure is adopted from Adelson (1995).

Chapter 2

Measuring Brightness Perception

Measuring human brightness perception cannot be directly measured by a physical tool, instead the measurements rely on experiments involving human participants. In general, experiments on brightness perception include visual stimuli such as images similar to the checkerboard in Figure 1.2. But instead of showing three-dimensional elements, in this study, the stimuli consist of two-dimensional elements. These elements are black and white surrounding context and gray areas inside the stimulus on the left and right, which are called *targets*. In Figure 2.1, typical stimuli from Domijan (2015) and Robinson, Hammon, and de Sa (2007) can be seen.

This work will focus on two particular methods to measure humam brightness perception: Fechner's (1860) *method of adjustment*, as utilized in the prior study by Bindermann (2022), and another method referred to as, *brightness ratings* introduced and used by Allaham (2022).

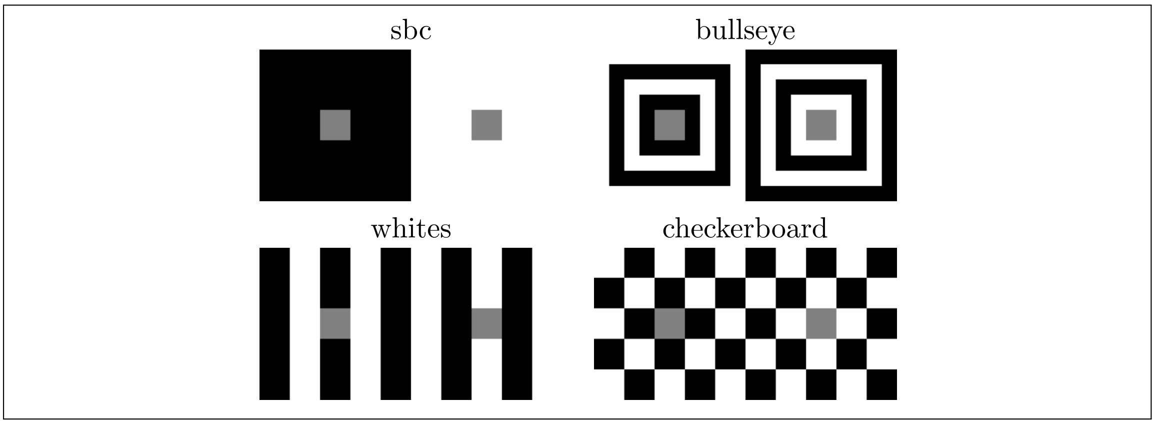


Figure 2.1: "sbc", "bullseye", "whites", and "checkerboard" are brightness illusions by Domijan (2015) and Robinson, Hammon, and de Sa (2007).

2.1 Brightness Ratings

In the brightness ratings experiment by Allaham (2022), participants were presented with stimuli, as one stimulus is shown in Figure 2.2. The participants were asked to rate the brightness of these targets in each stimulus using an ordinal five-point Likert-type scale. The scale consists of five options: ‘Left target is definitely brighter’, ‘Left target is maybe brighter’, ‘Targets are equally bright’, ‘Right target is maybe brighter’ and ‘Right target is definitely brighter’. At the end, brightness ratings gives the relative judgments from the participants.

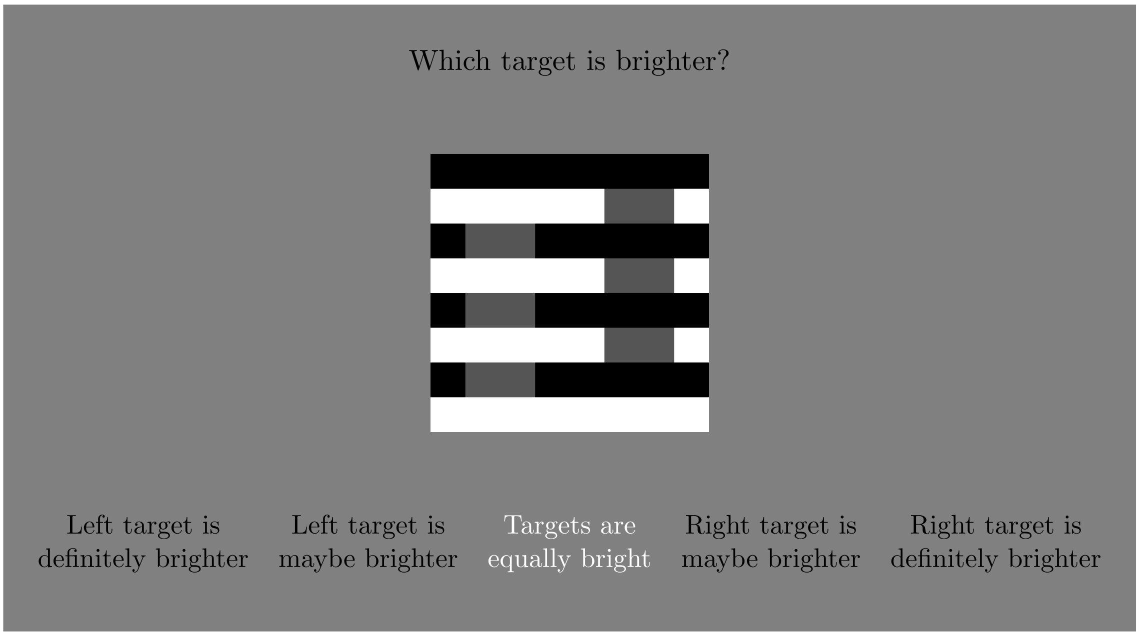


Figure 2.2: Experiment representation of the brightness ratings adopted from Allaham (2022). It shows the question at the top and the available predefined answer options at the bottom, requiring participants to select only one option whereas here the option ’Targets are equally bright’ is selected. In the middle of the Figure, there is the stimulus featuring two equiluminant targets positioned on the left and right side. The targets are surrounded by eight horizontal stripes alternating between black and white. In the middle left, three targets are positioned on the black stripes. Similarly, in the middle right, three targets are positioned on the white stripes, which places them each one stripe above the targets on the left. The question, stimulus, and answer options are presented on a gray background with a higher luminance than the equiluminant targets.

2.2 Method of Adjustment

In contrast to the brightness ratings task, the method of adjustment does not directly ask participants for their perceptual rating. Instead, participants were asked to adjust the physical luminance of a stimulus until it matches the target perceptually. Unlike in the brightness ratings method, the method of adjustment does not provide the relative judgments but rather absolute judgments.

In Figure 2.3, an experiment representation from Bindermann (2022) is shown. Participants were instructed to adjust the brightness gray square (2) at the top to match the brightness of the target (4) inside the stimulus (3).

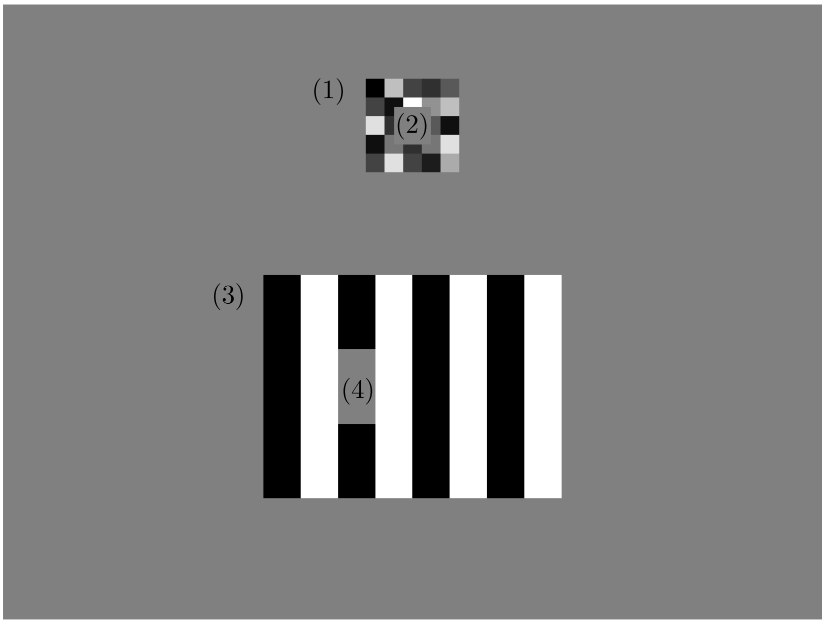


Figure 2.3: Experiment representation of the method of adjustment adopted from Bindermann (2022). It showcases an external comparison box at the top and the actual stimulus at the bottom. It comprises two gray areas: one in the top comparison box (2) and one in the bottom inside the stimulus (4). Both gray square areas are surrounded by contextual elements (1) and (3) presented on a gray background with the same luminance as the target (4). The surrounding context on top (1) consists of 24 randomly arranged small squares, effectively separating the gray area (2) from the background. On the other hand, the surrounding context (3) at the bottom is stimulus-dependent. The target inside the stimulus (4) is enclosed by eight alternating vertical stripes of black and white, specifically positioned vertically in the middle and on the second black stripe from the left where the first stripe from the left is black on this stimulus with an alternating pattern.

2.3 Comparing Methods

Both methods measure human perception for different visual stimuli without a time limit. Method of adjustment takes more time as it requires the participant to carefully and continuously adjust and compare the luminance of the target and the luminance of their adjustment. While in brightness ratings participants only have to compare both targets and select the most suitable predefined option for them.

Regarding the responses, the method of adjustment provides data that can be more amenable to quantitative analysis given its absolute values which are the adjusted luminance values by participants in candela per square meter (cd/m^2). In comparison, brightness ratings yield ordinal data, which might limit the range of statistical tests that can be applied. However, ordinal data might be more straightforward to interpret in terms of perceptual differences.

2.4 Thesis Objective

Given the results of the two methods, it is clear that a reliable method to collect data for brightness perception studies is essential. If the brightness illusions are consistently observed across different methods, it reinforces the robustness and generalizability of these brightness illusions. If results differ significantly between the methods, it raises new questions about why these discrepancies exist, paving the way for future research. Furthermore, understanding the nuances and potential differences between methods can inform the design of future experiments, leading to more accurate and comprehensive findings. With these considerations in mind, the research question of this study emerges: "Do data collected with two different methods - method of adjustment and brightness ratings - give similar estimates for the perceived brightness in different brightness illusions?". To address this question, a comparison of data collected in prior works employing both methods will be initiated in the next chapter.

Chapter 3

Data From Prior Works

After understanding the two methodologies employed in measuring brightness perception, the aim of this chapter is to look at the existing results from prior works using these methodologies. In the experiment by Allaham (2022), 10 of his 45 stimuli, shown in Figure 3.1, share some similarity to the 10 stimuli from the experiment by Bindermann (2022). But these stimuli from both experiments are not the same because some are stretched, twisted, flipped, have different sizes or luminances. From these stimuli the results can be compared as a first step.

3.1 Brightness Ratings

Allaham (2022) used brightness ratings in his research and measured the input from his 16 participants with 45 stimuli from Robinson et al. (2007), Domijan (2015), and Murray (2020). The responses from the participants are depicted as numbers representing their ratings, as shown on the x-axis in Figure 3.2. It also shows the *average direction of effect* that is determined by the median responses to the respective stimulus. A leftward direction of effect means the left target was perceived as brighter on average and vice versa for the rightward direction. While the stimuli "d_white", "r_WE_thick" and "r_sbc_large" have a leftward direction of effect, indicated by the color and position of the data points, all other stimuli have a rightward direction of effect.

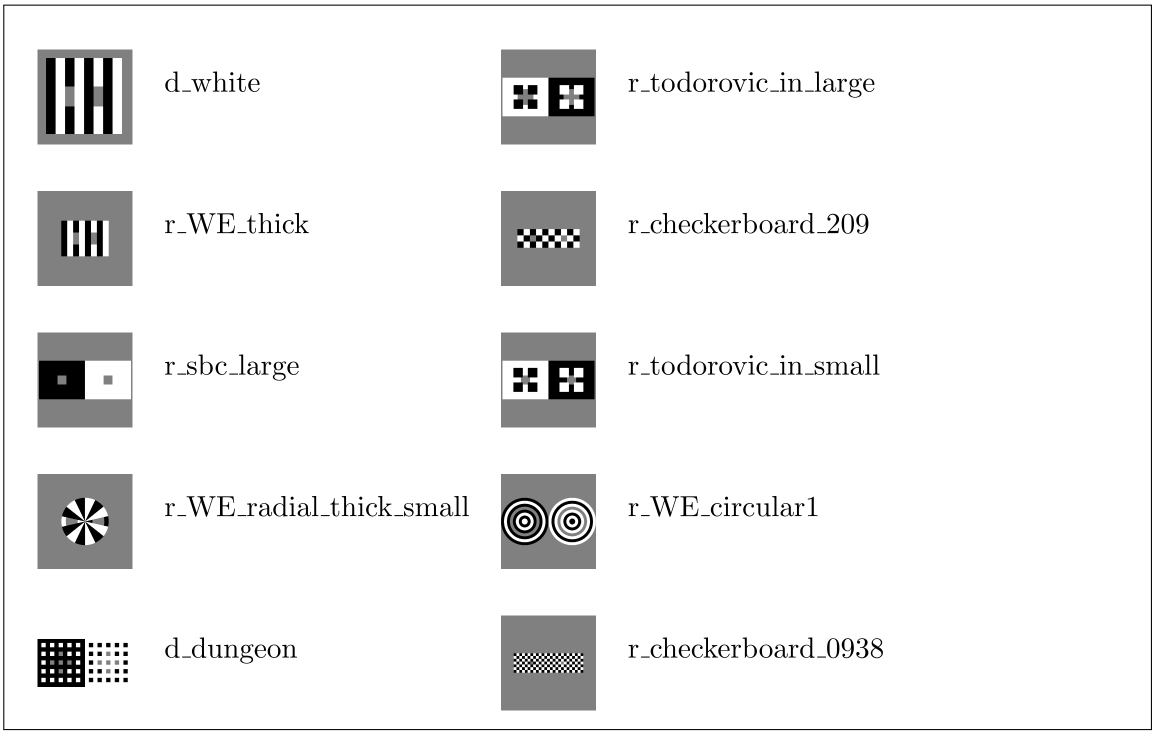


Figure 3.1: Subset of stimuli used in brightness ratings from prior study of Allaham (2022).

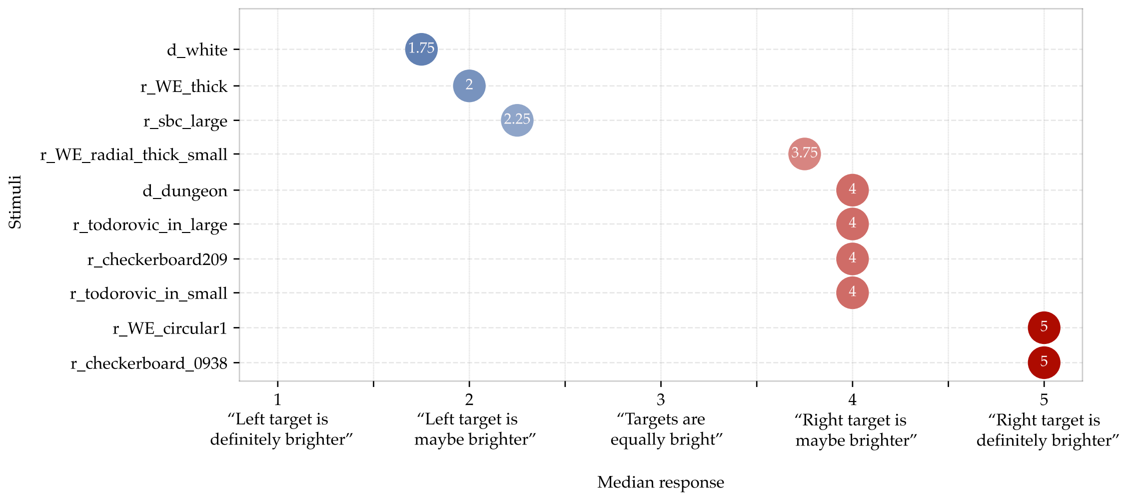


Figure 3.2: The median response for the 10 stimuli. The x-axis denotes the Likert-scale in numbers and the y-axis shows the brightness illusions presented to the participants. The distinction between the bluish and reddish data points indicates the difference of the average direction of each stimulus. Figure is adopted from Allaham (2022).

3.2 Method of Adjustment

In contrast, Bindermann (2022) used the method of adjustment to measure the brightness perception from his 27 participants and used 10 stimuli brightness illusions from Robinson et al. (2007) and Domijan (2015), as illustrated in Figure 3.3. The responses from the participants are their luminance adjustments for the targets. In Figure 3.3, higher adjustment values on the y-axis suggest a brighter perception for the corresponding target of the respective stimulus. Only the stimulus "White-groß" and stimulus "White-klein" have targets where the left target on average is perceived as brighter. For all other stimuli, the right or bottom target is perceived on average brighter.

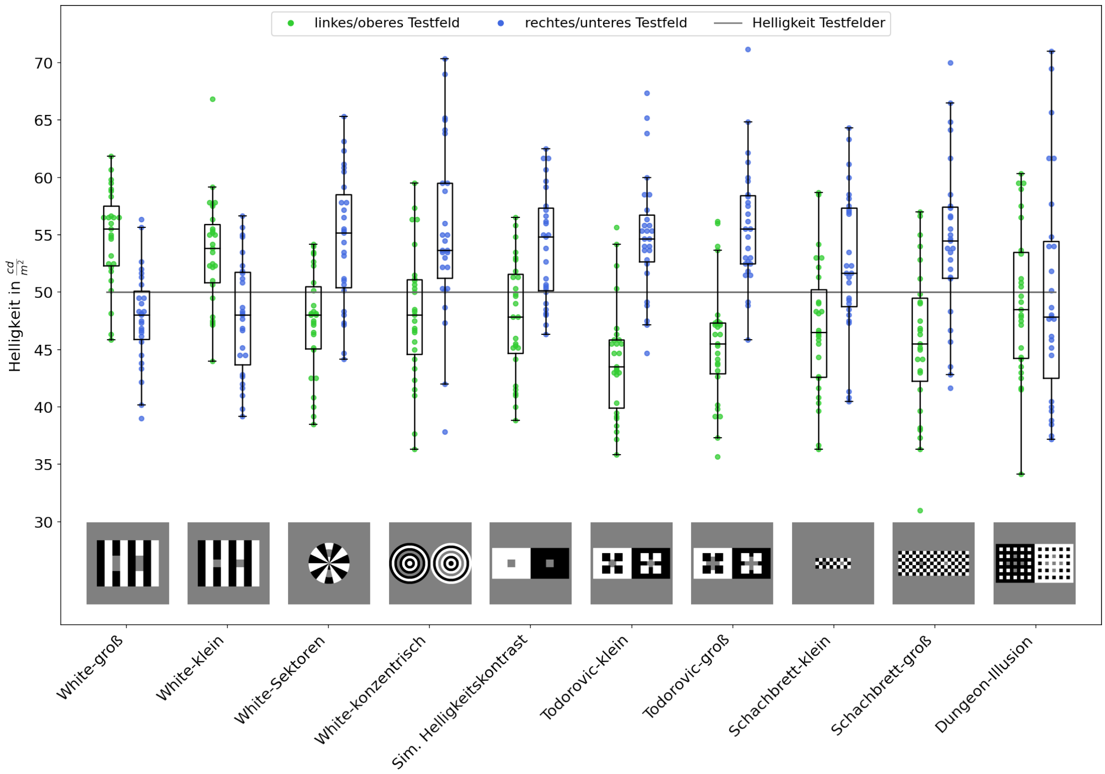


Figure 3.3: The arithmetic mean of participants' adjusted luminance values across all runs represented as data points. The x-axis denotes the brightness illusion. The y-axis signifies the luminance in cd/m² of the externally adjusted gray comparison box. The distinction between the green and blue data points indicates the difference in perception between the left/top and right/bottom target(s), respectively. The horizontal line serves as a benchmark, indicating the physical luminance level of the target. Figure is adopted from Bindermann (2022).

3.3 Comparing Data

One way to compare the results from brightness ratings and the method of adjustment is by looking at the participant's brighter perceived target on average, as shown in Table 3.1. In all comparison cases, both results from these two methods agree on the average direction of effect when the similar context surrounding the target is on the same side. For instance, the fifth comparison in Table 3.1 has a different direction of effect, but both agree that the target surrounded by black is perceived as brighter. This is because Allaham (2022) flipped his stimulus horizontally. A similar case can be seen in the third comparison, where he rotated his stimuli by 90 degrees counterclockwise.

3.4 Limitations

While the data comparison provides a glimpse into the similarities and differences between brightness ratings and the method of adjustment, it is imperative to recognize the limitations.

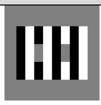
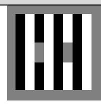
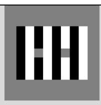
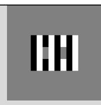
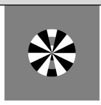
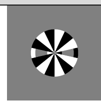
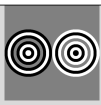
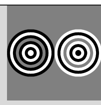
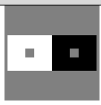
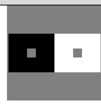
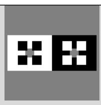
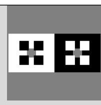
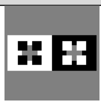
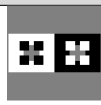
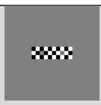
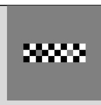
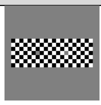
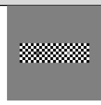
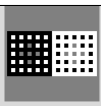
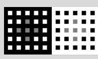
Firstly, Table 3.1 only shows the direction of which the target is perceived as brighter by adjusting or rating the stimuli. Bindermann's (2022) results, for instance, show that the "Dungeon-Illusion" stimulus produces different results from his other stimuli in terms of the intensity of the luminance adjustments. This is because the average adjusted luminance value for the left target in "Dungeon-Illusion" is 49.2 cd/m^2 and the average adjusted luminance value for the right target is 49.8 cd/m^2 making a difference of only 0.6 cd/m^2. The average difference of the targets from all of his stimuli is 6.96 cd/m^2. Consequently, it can be inferred that the "Dungeon-Illusion" stimulus was perceived with nearly equal brightness for both targets.

The second limitation lies in the participant pool. Since the data stems from two distinct groups of participants, individual biases and perceptual nuances are not controlled across methods. Therefore, any variance in results might be attributed to these individual differences rather than to the methodology itself.

Lastly, the stimuli presented in both methods were not identical. Variations in shape, size, and luminance values might have influenced the participants' perceptions. Consequently, the similarity in average results might stem from these variations, obscuring the true impact of the methodologies on perception.

In light of these observations, drawing a definitive conclusion from this comparison becomes a challenge. Yet, it lays the groundwork for formulating a hypothesis: "No significant difference exists between perceived brightness estimates derived from the method of adjustment compared to brightness ratings for a given brightness illusion." To test this hypothesis and answer the research question, an experiment using both methodologies on the same participant group becomes essential.

Table 3.1: Comparison of the direction of effect of stimuli in both methods from Bindermann (2022) and Allaham (2022).

	Bindermann's Experiment		Allaham's experiment	
Nr.	Stimuli	Average direction	Average direction	Stimuli
1	 White-groß	Left	Left	 d_white
2	 White-klein	Left	Left	 r_WE_thick
3	 White-sektoren	Bottom	Right	 r_WE_radial_thick_small
4	 White-konzentrisch	Right	Right	 r_WE_circular1
5	 Sim. Helligkeitskontrast	Right	Left	 r_sbc_large
6	 Todorovic-klein	Right	Right	 r_todorovic_in_small
7	 Todorovic-groß	Right	Right	 r_todorovic_in_large
8	 Schachbrett-klein	Right	Right	 r_checkerboard_209
9	 Schachbrett-groß	Right	Right	 r_checkerboard_0938
10	 Dungeon-Illusion	Right	Right	 d_dungeon

Chapter 4

Methodology

To answer the research question whether the results from brightness ratings and method of adjustment give similar results, data from prior studies collected with these methods has been compared. Due to the limitations of having different participant groups and different stimuli properties in the two prior studies, conducting an experiment addressing these limitations is essential. Therefore, the stimuli have been generated for both methods with the same piece of code to have the same properties such as size, shape, position and luminance. Besides, the same participants will conduct the experiment with both methods to resolve the limitations.

4.1 Stimuli

When choosing the stimuli, it is important to find an appropriate compromise between the acceptable amount of effort for the participants in the experiment and the most representative stimuli selection for the analysis. After calculating and predicting the effort based on the previous work, the number of stimuli is set to nine. The most representative stimuli brightness illusions are: (a) “sbc”, (b) “bullseye”, (c) “whites” by Domijan (2015) and (d) “checkerboard” by Robinson et al. (2007), as shown in Figure 4.1.

The other remaining stimuli are novel stimuli that serve as intermediate or transitional states between different distinct stimuli. They enable a transition from one stimulus to another and facilitate further analysis between the stimuli and methods. For instance, as shown in Figure 4.1, the novel brightness illusions (e) and (f) are transitions from (a) to (b).

This section will detail about the design of the stimuli, as shown in Figure 4.1. In this experiment, every stimulus consists of two equiluminant gray targets and their black and white surrounding contexts. All targets are square and have the same size, position, and luminance in each stimulus. The position of the targets is always on the left and right side within a stimulus, and there is white context above and below each right target, except for “strip”.

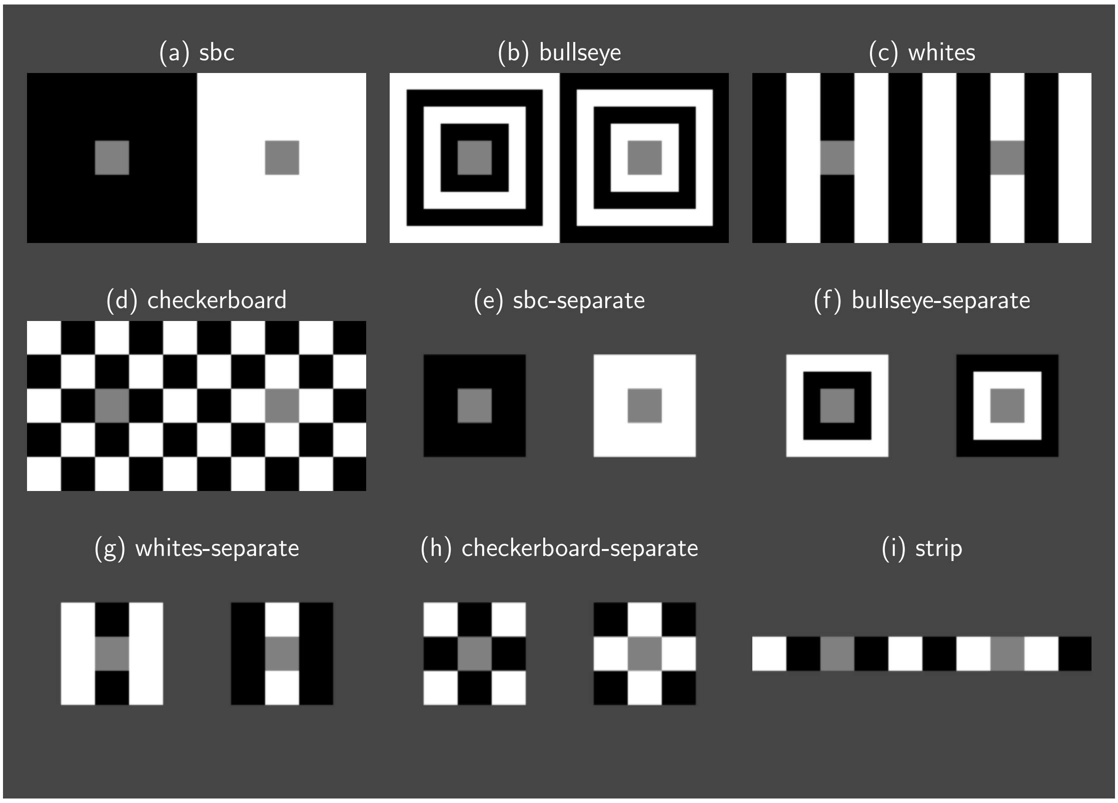


Figure 4.1: Nine stimuli that will be used in this experiment. (a), (b), (c), and (d) are original stimuli from Domijan (2015) and Robinson, Hammon, and de Sa (2007). (e), (f), (g), (h), and (i) are novel stimuli.

"Original" Stimuli

In stimulus (a) "sbc", the target on the left side is surrounded by black, while the target on the right is surrounded by white. Both frames on the left and right have the thickness of the double width or height of the targets.

In stimulus (b) "bullseye", the target on the left side is surrounded by a black square frame, which is then surrounded by a white square frame, which is again surrounded by a black square frame, which is finally surrounded by a white square frame. The target on the right is surrounded in a similar manner to the target on the left, but with the black and white reversed. Therefore, both targets are surrounded by four alternating black and white square frames. All frames on the left and right have a thickness of half the width or height of the targets.

In stimulus (c) "whites", alternating five black and five white vertical stripes can be seen. Starting from the left with the black vertical stripe and ending on the left with a white stripe. On the third stripe from the left which is a black vertical stripe, in the middle of the stripe there is the target which has the width and height of the thickness of the stripe. On the third stripe from the right, which is a white vertical stripe, in the middle of the stripe, there is the target that has the width and height of the thickness of the stripe.

In stimulus (d) “checkerboard”, there is a two-dimensional checkerboard consisting of black and white tiles which has all the same size as the two targets. The “checkerboard” has 5 rows and 10 columns. On the top left and bottom left, there is a white tile, and on the top right and bottom right, there is a black tile. The left target is positioned on the third row from the top and third column from the left. The right target is positioned on the third row and third column from the right.

Novel Stimuli

After detailing the the original stimuli, the novel stimuli will be discussed.

In stimulus (e) “sbc-separate”, the target on the left is surrounded by a black square frame, which is surrounded by a gray background which has a lower luminance than the targets. The target on the right is surrounded by a white square frame, which also is surrounded by a gray background that also surrounds the black square frame on the left. Both frames on the left and right have the same thickness as the width and height of the targets.

Stimulus (f) “bullseye-separate” is like stimulus (b) “bullseye” but the last two outer frames on both sides are removed and replaced with a gray background that has a lower luminance than the targets.

In stimulus (g) “whites-separate”, the target on the left is surrounded left and right by two white vertical stripes, one on the left and one on the right. These stripes have a height three times that of the target’s height. Also, the target on the left is surrounded by two black squares, one above and one below. All these elements that surround the left target are surrounded by a gray background that has a lower luminance than the targets. On the right we have the same as on the left but black and white are swapped. This is also surrounded by the same background that surrounds the elements on the left.

Stimulus (h) “checkerboard-separate” is like stimulus (d) “checkerboard” but the first and last row are removed. The first, fifth, sixth and last column are also removed. All the removed elements are being replaced with a gray background that has a lower luminance than the targets.

Stimulus (i) “strip” is like stimulus (d) “checkerboard” but every row except the third is removed resulting in a “strip” in the middle that is surrounded above and below by a gray background. This stimulus is also the reason that the background has a lower luminance than the targets, in order to separate the targets from the background.

The gray background surrounding the white and black context has a luminance of 75 cd/m^2 which is lower than the target’s luminance of either 122.5, 125 or 127.5 cd/m^2. Although there are variations in the target’s luminance for each stimulus, which will be explained in the next section (4.2), both targets on the left and right of one stimulus always have the same luminance.

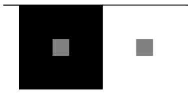
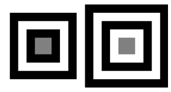
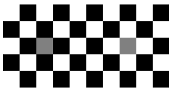
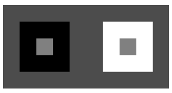
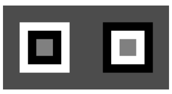
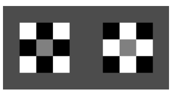
Building on the stimuli presented, this section explains the implicit brightness effect of the stimuli based on the prior works of Allaham (2022) and Bindermann (2022). In certain scenarios, for instance in the “sbc” stimulus, a target surrounded by only a dark area is perceived as brighter, while a target surrounded by only a bright area is perceived as darker.

However, when the stimulus presents a clear pattern, humans typically compare targets to the pattern rather than their immediate surroundings. For example, in a sequence alternating between black and white, if a target is placed in a position where, according to the sequence, black should appear, humans compare the target to the expected black pattern. Stimulus (b) "bullseye" is a suitable example for demonstrating this concept. It reveals that the right target, despite being encircled by white, seems brighter than the left target surrounded by black. This perception arises due to the alternating black and white pattern present in the stimulus.

In the novel stimuli, there are no clear expectations as to which target would be perceived as brighter because no prior work has been done on these stimuli. It can only be hypothesized that they have a similar brightness effect expectation as the expectations from their "unseparated" versions.

Applying all these brightness effects results in Table 4.1.

Table 4.1: The hypothesized brightness effects for the stimuli used in the experiment.

Stimulus	Stimulus name	Hypothesized brighter target
	"sbc"	Left
	"bullseye"	Right
	"whites"	Left
	"checkerboard"	Right
	"sbc-separate"	Left
	"bullseye-separate"	Right
	"whites-separate"	Left
	"checkerboard-separate"	Right
	"strip"	Right

4.2 Design

Next, the focus shifts to the methodology design, particularly how the stimuli are presented and integrated into the experiment, to understand the methods employed.

In the brightness ratings experiment all targets of all stimuli had the same size, shape, and position. This made the transition from one trial to the next seem to look like the targets stayed on the screen and only the surroundings changed. This could lead to the impression that the targets were the same targets from the previous trial, resulting in the same response as the previous response. To fix this problem, in both methods, an inter-stimulus interval has been implemented between each trial. In addition, a variation in the target's luminance has been implemented within the range of 122.5, 125, and 127.5 cd/m^2. This problem did not exist in the prior work because either the size, shape, or position of the targets was different for many stimuli.

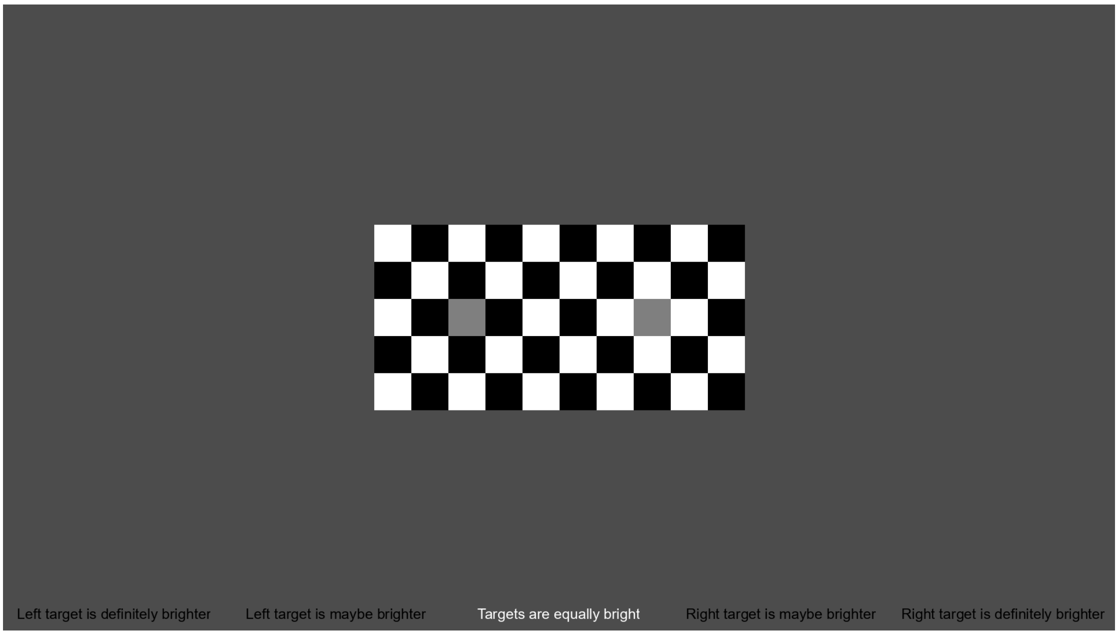


Figure 4.2: Screenshot from a trial showing the checkerboard stimulus in the method of adjustment experiment.

In the method of adjustment, to facilitate participants' identification of the reference for adjusting the comparison box, one target was excluded from the stimulus. Thus, Bindermann's stimuli had only one target, while Allaham's stimuli showed both targets. As this is part of the method of identifying the target's side, it will not be changed. An additional small horizontal black bar has been placed above either the left or right target and stimulus to make identification of the target easier, as shown in Figure 4.3.

For the brightness ratings, each brightness illusion was presented in six trials in one session. In three trials, the brightness illusion was presented at three different target's luminance, and in each of these trials, a horizontally flipped version was also presented, for a total of six trials. In the method of adjustment, each brightness illusion is also presented in six trials. In three trials, the brightness illusion is presented at three different

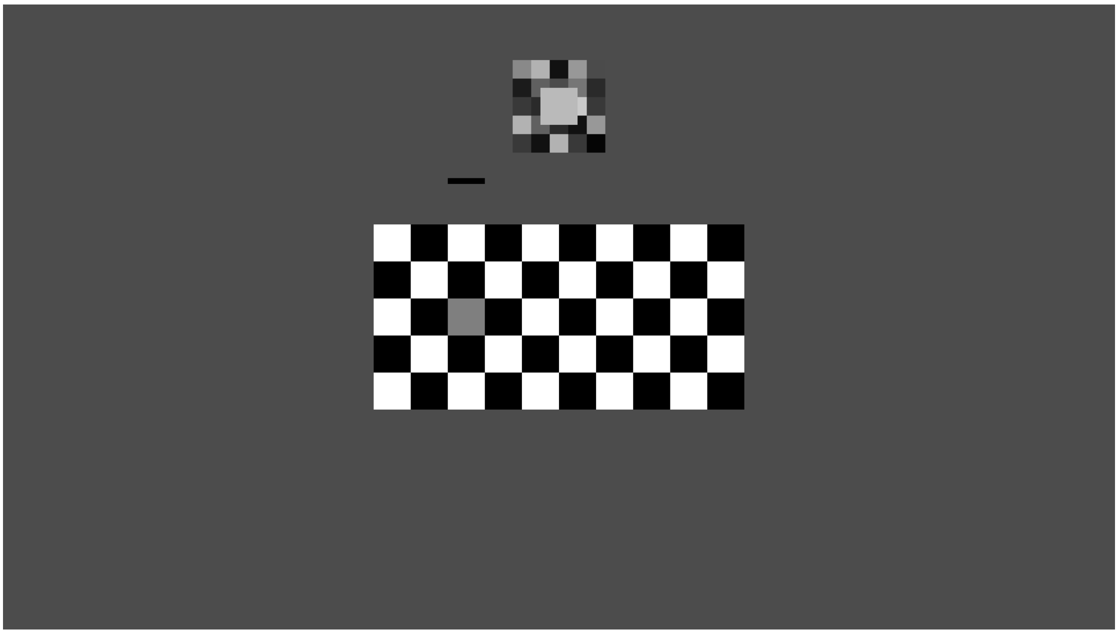


Figure 4.3: Screenshot from a trial showing the checkerboard stimulus in the brightness ratings experiment.

target's luminance, and each of them presents a version showing only the right and left target, for a total of six trials. Again, the stimuli were also flipped in the method of adjustment, but randomly, so as not to increase the number of trials and make it unbearable for the participants. Six trials for all nine brightness illusions result in 54 trials in each method.

In addition to the stimuli, catch trials, as shown in Figure 4.4, were implemented in both methods to ensure that participants paid attention throughout the experiment and performed the tasks according to the instructions. It also serves to prevent participants from not choosing an extreme end of a scale, which is the case in many rating experiments according to Cunningham and Wallraven (2011). There are 30 catch trials evenly distributed between the main trials in the brightness ratings method and 12 catch trials in the method of adjustment. These 42 catch trials presented the catch stimuli or a horizontally flipped version of these catch stimuli. Their target luminance and their usage in the the methods as listed in Table 4.2. There is an objective, physical difference between the targets in many catch stimuli, as shown in Table 4.2, so the participant should give a "correct" answer. If the responses of a participant to the stimuli differ strikingly from the average, the responses to the catch trials can be reviewed. Based on this, three participants have been excluded from the analysis because they did not complete the task according to the instructions. To maintain a consistent comparison basis, data from participants excluded from one method were also excluded from the other method. This ensures uniformity in data treatment across the study, maintaining the integrity of the results. The responses of one participant in brightness ratings has been inverted since the results of the catch trials show a clear sign that this participant inverted the task, prompting them to determine which target appeared darker. 54 trials in each method and a total of 42 catch trials make it a total of 150 trials.

There is an objective, physical difference between the targets in many catch trials, as shown in Table 4.2, thus the participant should give a "correct" answer.

Table 4.2: Target's luminance of the catch stimuli

Catch stimulus	Target luminance (cd/m^2)		Usage in method	
	Left	Right	Brightness ratings	Method of adjustment
"catch-stimulus-black-1"	150	100	✓	x
"catch-stimulus-black-2"	137.5	112.5	✓	x
"catch-stimulus-black-3"	125	125	✓	✓
"catch-stimulus-black-4"	112.5	137.5	✓	x
"catch-stimulus-black-5"	100	150	✓	x
"catch-stimulus-white-1"	150	100	✓	x
"catch-stimulus-white-2"	137.5	112.5	✓	x
"catch-stimulus-white-3"	125	125	✓	✓
"catch-stimulus-white-4"	112.5	137.5	✓	x
"catch-stimulus-white-5"	100	150	✓	x

4.3 Setup

The experiment took place in a laboratory of the Computational Psychology department. It was equipped with a computer, monitor, controller, chin and forehead rest. The monitor, controller, chin, and forehead rest were on a desk with a fixed distance of 80 cm from forehead to the monitor. Besides all that, there was also a chair where a participant sat and put their chin, and forehead on the chin and forehead rest to do the experiment. All this equipment was housed in a controlled light environment, enclosed by a thick black curtain from the top of the ceiling down to the floor, ensuring no light, except from the monitor, was emitted. From outside, the activity of the participants could be monitored from another monitor connected to the same computer.

The monitor that is being used to present the stimuli to the participants is a *VIEWPixx/3D LCD* with a resolution of 1920 x 1080 pixels and a refresh rate of 120 Hz. This 24-inch monitor with 16-bit resolution is able to display luminance values ranging from 0 to 250 cd/m^2. The controller which is our input device is a *RESPONSEPixx* response box with 5 buttons and connected to the computer. The chin and forehead rest is an *SR Research Head Support* that can be adjusted in height.

The experiment software was programmed in Python, following the standardized template provided by the research group for laboratory experiments. The Python library 'stimupy'[1] and 'hrl'[2], which were developed and provided by the research group of Computational Psychology, have been utilized in this experiment.

[1] https://github.com/computational-psychology/stimupy
[2] https://github.com/computational-psychology/hrl

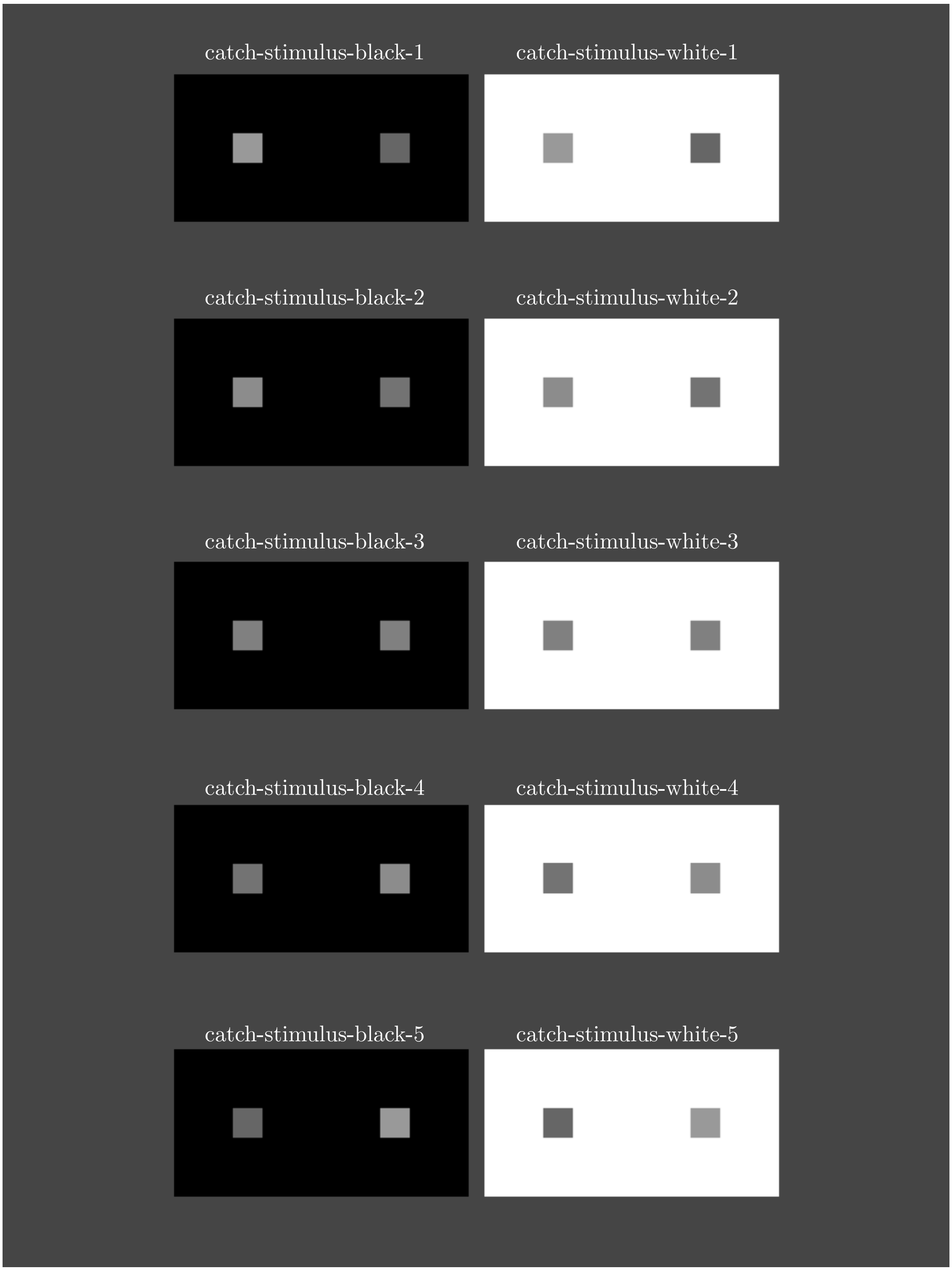


Figure 4.4: Each catch stimulus have either a black or white background. All these stimuli was shown in the brightness ratings, whereas in the method of adjustment only catch stimuli with equiluminant targets, which is “catch-stimulus-black-3” and “catch-stimulus-white-3”, were used and only one target was shown. The targets’ luminance values are listed in Table 4.2.

4.4 Procedure

Upon recruiting potential participants for the study, appointments were set up with those expressing interest. These were members of the Computational Psychology department, students of the Technical University of Berlin, people from the author's personal environment, and the author himself. When participants arrived, they were informed about the experiment's objectives and procedures. The method of adjustment and brightness ratings were conducted in a single session. The order of the method was randomized. Randomizing variables in psychophysical experiments helps minimize bias and increase generalizability of the results, as noted by Goodwin and Goodwin (2016). Throughout this session, participants had the autonomy to take breaks and make inquiries as they saw fit. Once all 150 trials had concluded, participants received a debriefing. Finally, participants received their compensation.

In this study, there were 21 participants, including the author. Each participant completed 150 trials, resulting in a total of 3,150 trials. On average, each session lasted about 28 minutes. This comprised roughly 5 minutes for instructions, formalities, and debriefing, 8 minutes for the brightness ratings experiment, and 15 minutes for the method of adjustment experiment.

To conduct this experiment, obtaining approval from the ethics committee was necessary. Additionally, every participant has to sign a declaration of consent before partaking in the experiment. These documents can be found in the appendix.

4.5 Data Questions

Data questions are defined to help answer the research question. The first data question is about the average direction of effect. Some stimuli may show a consistent direction of effect, whether it be leftward direction if the left target is perceived as brighter, or rightward direction if the right target is perceived as brighter. Identifying patterns among stimuli in terms of the average direction becomes paramount.

1^{st} DATA QUESTION: Do the data collected in both methods give the same average direction of effect?

The second data question is about the strength of effects. It is introduced to measure the strength of a direction of effect within a direction. In the method of adjustment, the intensity is measured, whereas in brightness ratings, the certainty regarding one target being brighter than the other target is measured.

2^{nd} DATA QUESTION: Do the data collected in both methods give similar strength of effect?

Lastly, the third question is about variability of brightness effects. Perceived brightness varies across stimuli. The aim is to detect if certain stimuli consistently cause striking variability in brightness perception among participants. The variability is assessed using the interquartile range of the adjustment differences of left and right targets. A higher interquartile range indicates higher variability. It is important to keep in mind that both methods use different metrics, so the extremes in one method and the other method are different, as well as the distances on the scale. Therefore, the interquartile range cannot be directly compared between these methods. Instead, the highest and lowest variability and striking patterns within each individual method will be examined.

3^{th} DATA QUESTION: Do the data collected in both methods give similar variability of effect?

Chapter 5

Results

Following the complete account of the methodology, the next chapter unveils the study's outcomes and answers all three data questions based on the collected data with both methods. The data from three participants were excluded because they failed to do the task according to the instructions.

5.1 Average Direction of Brightness Effects

5.1.1 Brightness Ratings

The first data question is about the average direction of effect. In the brightness ratings task, participants directly indicated the direction of effect. To get the average direction of effect, the responses have been converted to numerical values and calculated the median, as this is the most appropriate measure of central tendency given the ordinal data. As shown in Figure 5.1, for the stimuli "sbc", "whites" and "whites-separate" left targets were perceived as brighter on average, while "sbc-separate" and "bullseye-separate" have targets that were perceived as equally bright resulting in having no average direction of effect. All other stimuli have right targets that, on average, were perceived as brighter. These findings for the four original stimuli go in the expected direction confirming the hypothesis.

For the novel stimuli it was expected that they go to the same direction as their original stimuli, which is true for "whites-separate" and "checkerboard-separate". "sbc-separate" and "bullseye-separate" do not have an average direction of effect. For "strip", it shows that it goes in the same direction as the "checkerboard" stimulus. This confirms the hypothesis as "strip" a part of the "checkerboard" but showing only the middle row of "checkerboard".

In Figure 5.7, the reason can be found for the two stimuli having no average direction of effect. There were a handful of participants, P4, P7, P8, P11, and P13, who on average responded that they perceived both targets as equally bright for the stimuli "sbc-separate" and "bullseye-separate". If the mean was taken to measure the average direction of effect, then these two stimuli would have the same direction of effect as their "original" stimuli where the target's surroundings are not separated. This shows how the median and mean give different results, as the mean assume equal intervals, for instance between -2 and -1 is the same as between -1 and 0 which is not the case in brightness ratings since the numbers are just representing an option.

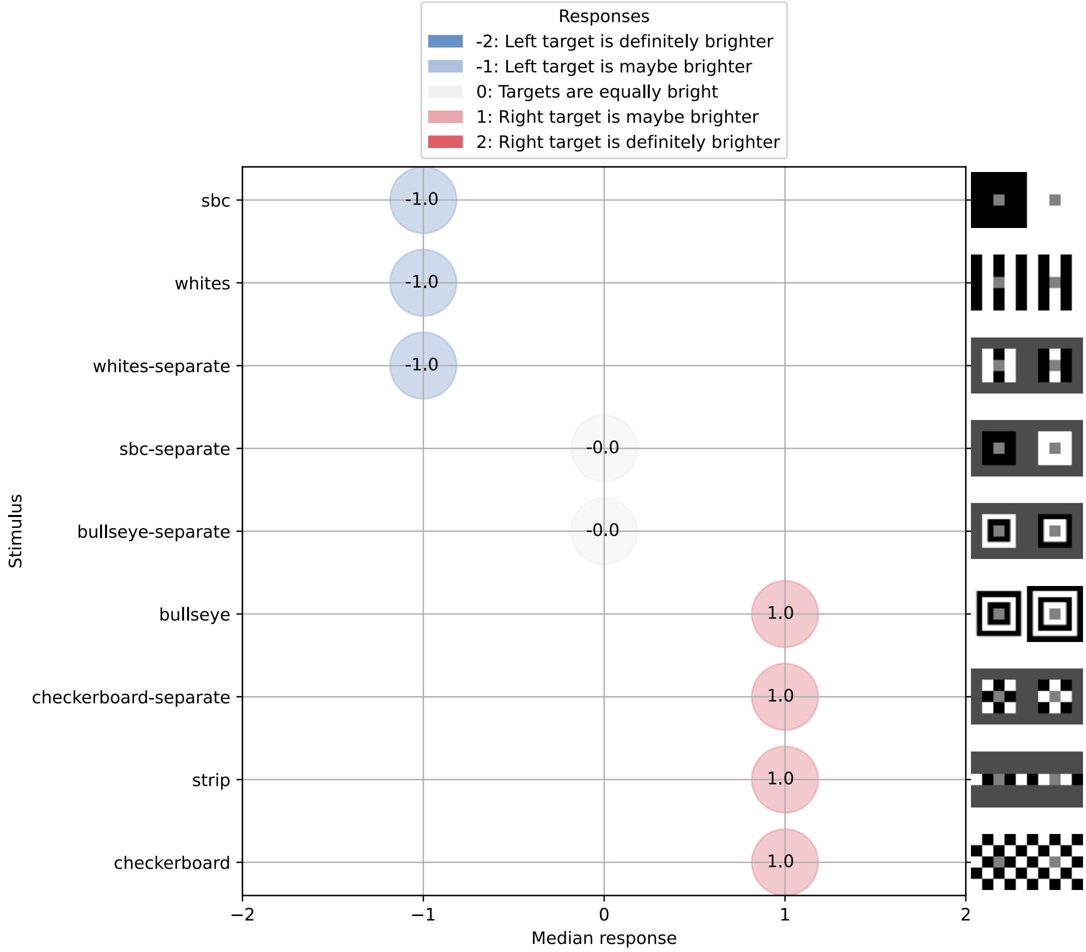


Figure 5.1: Brightness ratings: median response from all participants to each stimulus. The y-axis delineates the stimuli. The x-axis at the bottom shows response value. Negative average responses are colored bluish and indicate a leftward direction of effect, while positive average responses are colored reddish and indicate a rightward direction of effect. Gray data points indicate no direction of effect.

5.1.2 Method of Adjustment

In the method of adjustment, the indicator for the direction of effect is derived from the difference between the adjustments for the left and right targets. By calculating the mean of all such indicators, one can determine the average direction of the effect, as shown in Figure 5.2. In other words, a target with a higher adjustment is the brighter perceived target.

For “checkerboard” and “strip,” the right target is perceived as brighter on average. All other stimuli, in the method of adjustment, have left targets that are perceived on average brighter than the right targets. It is worth mentioning that “checkerboard-separate”, “bullseye” and“strip” have very low difference of under 4 cd/m^2. This means they are very close to zero, implying it has a direction of effect that is close to non-existent. Every result from each stimulus confirms the hypothesis except the result from “bullseye”, “bullseye-separate” and “checkerboard-separate”. All these average direction of effect results can be supported by the individual adjusted luminance from each participant and their direction, showing there is no major outliers that skewed the whole result, as shown in Figure 5.3.

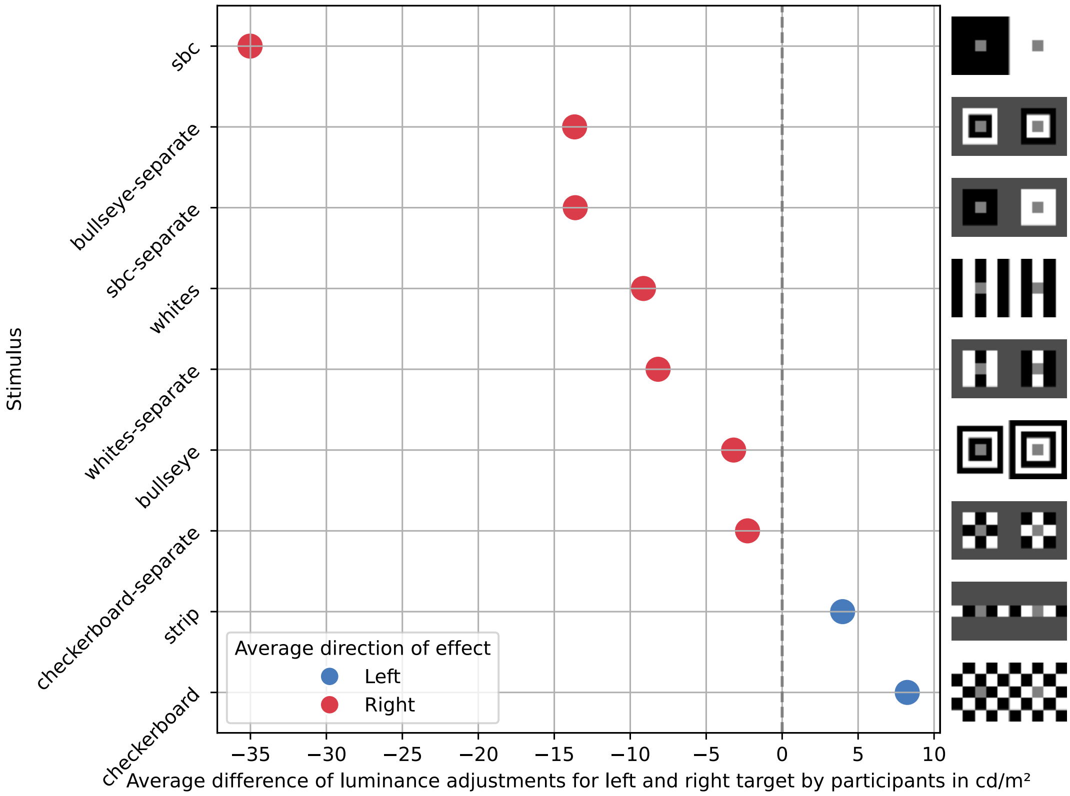


Figure 5.2: Method of adjustment: Average (mean) differences of the adjusted luminance to match the corresponding target. The y-axis lists stimuli with their left and right target in ascending order, based on the difference of the luminance adjustment between the targets. The x-axis scales the luminance in cd/m^2. Blue data points represent the average luminance adjusted to match the left target, while red dots signify the same for the right target. The dashed horizontal lines represent the actual luminances of the target that were presented to the participants during the experiment trial.

5.1.3 Comparing Average Direction of Effects

Five out of the nine stimuli have the same average direction of effect in both methods, as shown in Table 5.1. This includes the three stimuli “whites”, “sbc” and “checkerboard”, as well as two novel stimuli “whites-separate” and“strip”. The rest of the stimuli have different results, where in the method of adjustment all have on average a leftward direction of effect and contrary in brightness ratings, “checkerboard-separate” and “bullseye” have on average a rightward direction of effect while “bullseye-separate” and “sbc-separate” have no direction of effect. The result of having no direction of effect in the method of adjustment would only appear if both average adjustments for left and right target have the exact same value.

From all stimuli, only “bullseye” in the method of adjustment has a result that is different from what is expected in the hypothesis. Upon further analysis, it is evident

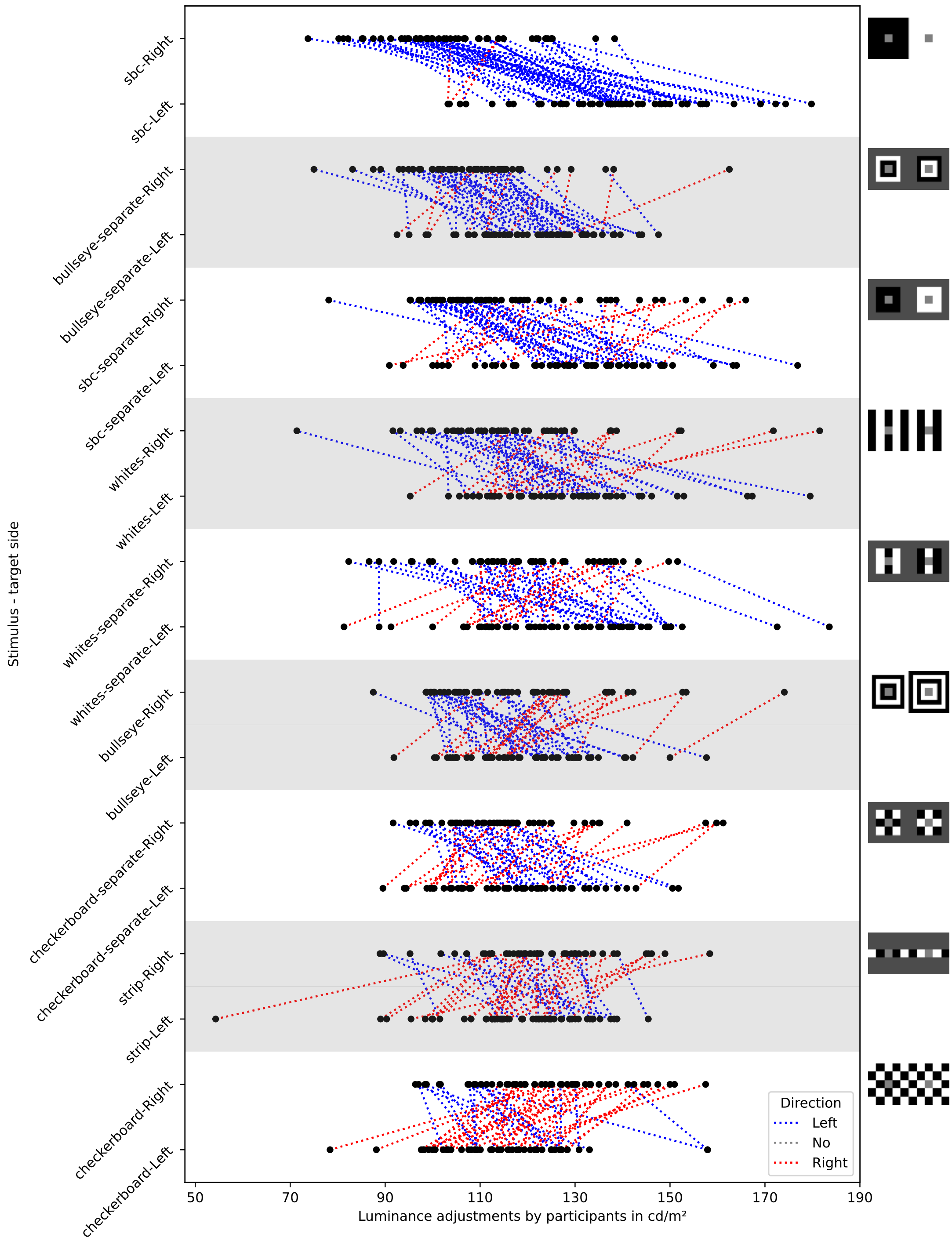
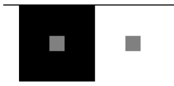
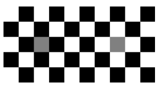
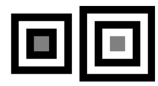
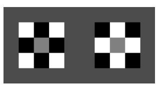
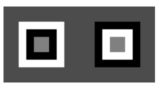
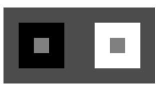


Figure 5.3: Method of adjustment: Each pair of data points connected by a dashed line represents a participant's adjustment. The x-axis displays luminance in cd/m². The y-axis shows stimuli in ascending order based on luminance difference between left and right targets.

that for this stimulus, the average direction of effect from individual participants is split between left and right. Specifically, as shown in Figure 5.4, half of the participants have a leftward average direction of effect while the remaining half, which carries more weight in the overall results, has on average a rightward direction of effect.

Another observation is that stimuli with targets encircled completely by only black or white which are “sbc”, “sbc-separate”, “bullseye-separate” and “bullseye”, always have the same average direction of effect in the method of adjustment.

Table 5.1: Comparison of average direction of effect for the respective stimuli in both methods.

Stimulus		Average direction of effect	
		Brightness ratings	Method of adjustment
	“sbc”	Left	Left
	“whites”	Left	Left
	“whites-separate”	Left	Left
	“strip”	Right	Right
	“checkerboard”	Right	Right
	“bullseye”	Right	Left
	“checkerboard-separate”	Right	Left
	“bullseye-separate”	No	Left
	“sbc-separate”	No	Left

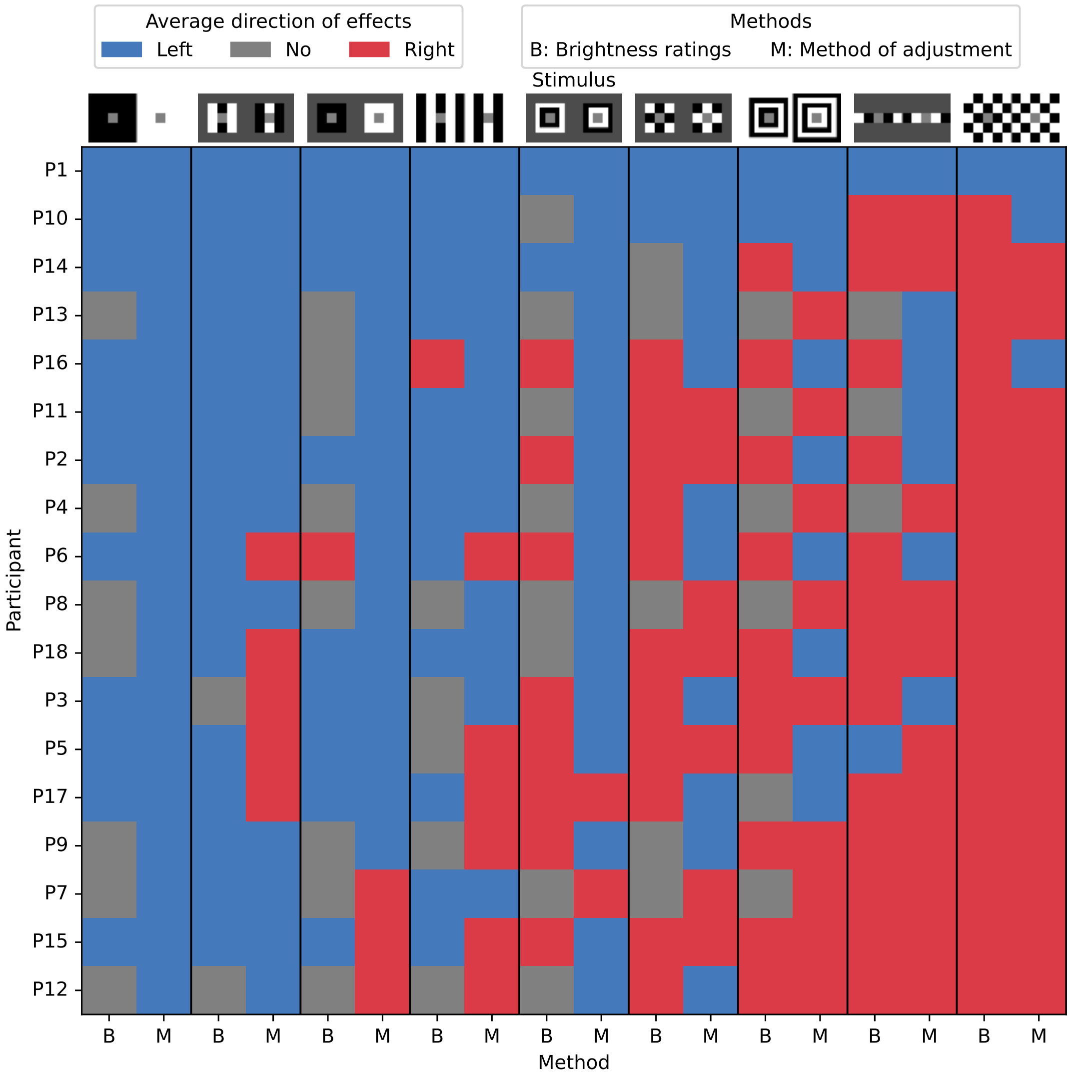


Figure 5.4: Comparison of the average direction of effect from each participant for each stimulus in the brightness ratings and method of adjustment. In brightness ratings, the median of the responses was taken to determine the average direction from each participant, while in the method of adjustment, the mean of the difference from the target's adjustments was taken. A negative number is presented by a bluish color and positive number by reddish color. Zero is presented by a gray color. The x-axis presents the stimuli and the methods, while the y-axis lists the participants. The y-axis lists all participants.

5.2 Strength of Brightness Effects

5.2.1 Method of Adjustment

The second data questions is about the strength of effect. In the method of adjustment, the strength of effect is indicated by the absolute difference of the average luminance adjustments between the left and right targets. A big absolute difference suggests a great strength of effect, as one target is adjusted noticeably brighter or darker than the other. Conversely, no difference means participants adjusted both targets equally bright, leading to a weak strength of effect.

The absolute differences from Figure 5.2 are shown in Figure 5.5 for a better overview. These data points can be ranked from the biggest to the the lowest adjustment difference, as Table 5.2 illustrates. An alternative method of representation involves segregating the data into two groups, 'weak' and 'strong' strength of effect, based on the boundary defined by the biggest gap in the sorted dataset. However, this method potentially obscures the nuances of data variability.

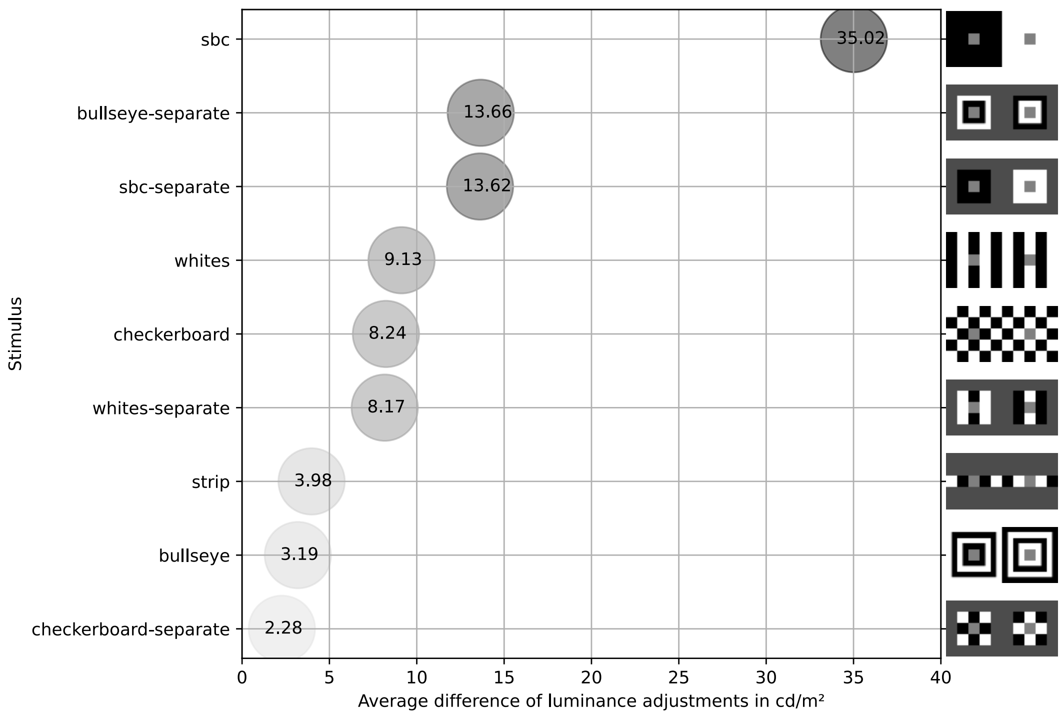


Figure 5.5: Methods of adjustment: Average absolute difference between the adjusted luminance for left and right targets to stimuli. The x-axis at the bottom displays luminance values. The y-axis depicts the stimuli, arranged in ascending order. The darker the data point, the greater the strength of effect.

5.2.2 Brightness Ratings

In contrast to the method of adjustment, brightness ratings has a different definition for the strength of effect. Here, a stimulus with a great strength of effect is assumed to produce a more certain response, which is -2 or 2. This implication does not work the other way around. A certain response might simply reflect a participant's general confidence, as shown in Figure 5.7 for the participants P1 and P15. According to Figure 5.1, there is no median response with a certainty of -2 or 2. This observation suggests that stimuli with any direction of effect have a greater strength of effect than stimuli with no direction of effect, as shown in Table 5.2. Even if the mode is taken as the measure for certainty, the result for the strength of effect would not change, as shown in Figure 5.6. The mode was not taken as the measure because it could miss subtler shifts in the data since the median provides a more continuous and representative measure of central tendency. Furthermore, this ensures consistency in the metrics that are being used in brightness ratings for the average direction of effect and strength of effect.

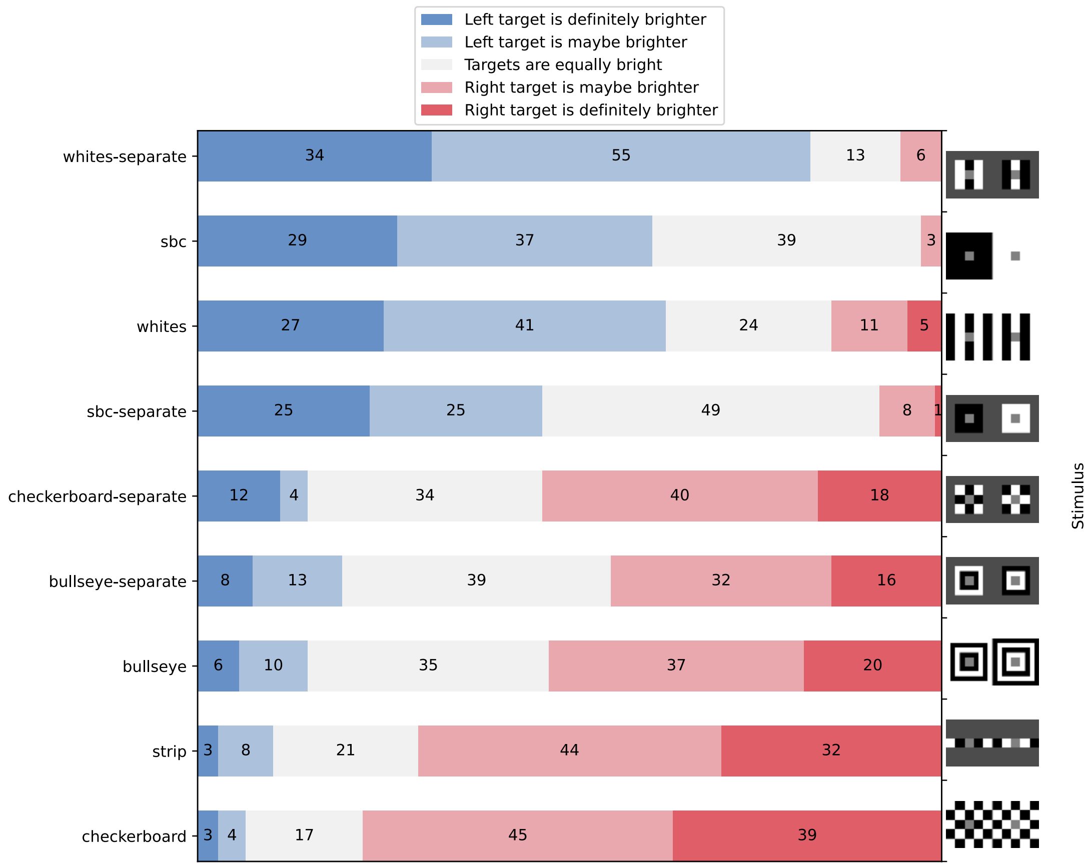
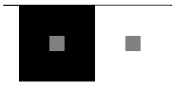
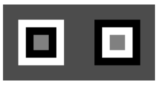
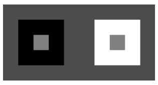
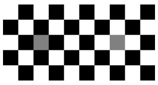
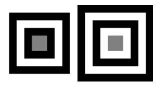
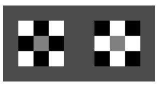


Figure 5.6: Brightness ratings: Stacked bar chart displaying the distribution of responses for each stimulus. The y-axis represent the stimuli. On each color-coded bar, we see the numbers of responses the stimulus received.

5.2.3 Comparing Strength of Brightness Effects

Rather than directly comparing the results, the rankings for the stimuli regarding the strength of effect from both methods are compared. The reason for this comparison is because the methods apply distinct metrics with varying intervals and limits. Table 5.2 reveals that only “sbc” along with “bullseye-separate” are the only stimuli that have the same ranking for their strength of effect in both methods. The rankings for all other stimuli differ. It is critical to note that the method of adjustment assigns a unique ranking to each stimulus, demonstrating greater variability. In contrast, the brightness ratings yield only two rankings. This information underscores the method of adjustment’s sensitivity in detecting variations among stimuli.

Table 5.2: The strength of effect ranking for the respective stimuli in each method.

Stimulus		Strength of effect: Certainty ranking in brightness ratings	Strength of effect: Adjustment difference ranking in method of adjustment
	“sbc”	1	1
	“bullseye-separate”	2	2
	“sbc-separate”	2	3
	“whites”	1	4
	“checkerboard”	1	5
	“whites-separate”	1	6
	“strip”	1	7
	“bullseye”	1	8
	“checkerboard-separate”	1	9

5.3 Variability of Brightness Effects

5.3.1 Brightness Ratings

As variability was already mentioned in the previous section, this section will closely examine variability within each method and answers the third data question about variability of effect. This is done by measuring the interquartile range of the results. In brightness ratings, most stimuli have an interquartile range of 1. Only “sbc” and“strip” have the highest variability with an interquartile range of 2, followed by “whites” with 1.25, as shown in Table 5.3

Upon analyzing individual responses, remarkable findings in Figure 5.7 show that participant P1 and P15 show a higher confidence than the average. When reviewing the raw data, participant P1 almost only perceived the left target as brighter except in one trial where the “checkerboard” stimulus with target intensity of 127.5 cd/m^2 was presented, this participant perceived the right target as maybe brighter. Every tile in Figure 5.7 represent the median of two responses to the same brightness illusion with the same target luminance. Participant P1 responded with -2 and 1 to the “checkerboard” (127.5 cd/m^2 target luminance) stimulus, leading to the median of -0.5. This participant was not discarded because of the high rate of exactly correct catch trial response.

There are also participants, such as P14, and P9, that deviate from the average response for multiple stimuli. For instance P14 have on average a leftward direction of effect for “bullseye-separate”, while P9 have on average a rightward direction of effect for the stimulus “sbc-separate”. Theses are the opposite direction of effect from the average. However, these participants maintain consistency for one direction and only differ on the other average direction of effect.

5.3.2 Method of Adjustment

In contrast to brightness ratings, in the method of adjustment the variabiltiy is derived by taking from each participant the difference of their adjustments for the left and right target to each stimulus with each respective target’s luminance. Then for all these differences, as shown in Figure 5.8, the interquartile range is calculated for each stimulus. A higher interquartile range indicates greater variability.

The method of adjustment offers a broader range of variability, ranging from“strip” with the lowest variability of 22.27 cd/m^2 to “sbc-separate” with the highest variability of 35.45 cd/m^2 as shown in Table 5.3.

Furthermore, Figure 5.8 support the interquartile range numbers and provides visual insight into the variability of the difference of each adjustment for the left and right target. Data points that are extensively spread out, signify higher variability, while densely clustered data points denote lower variability.

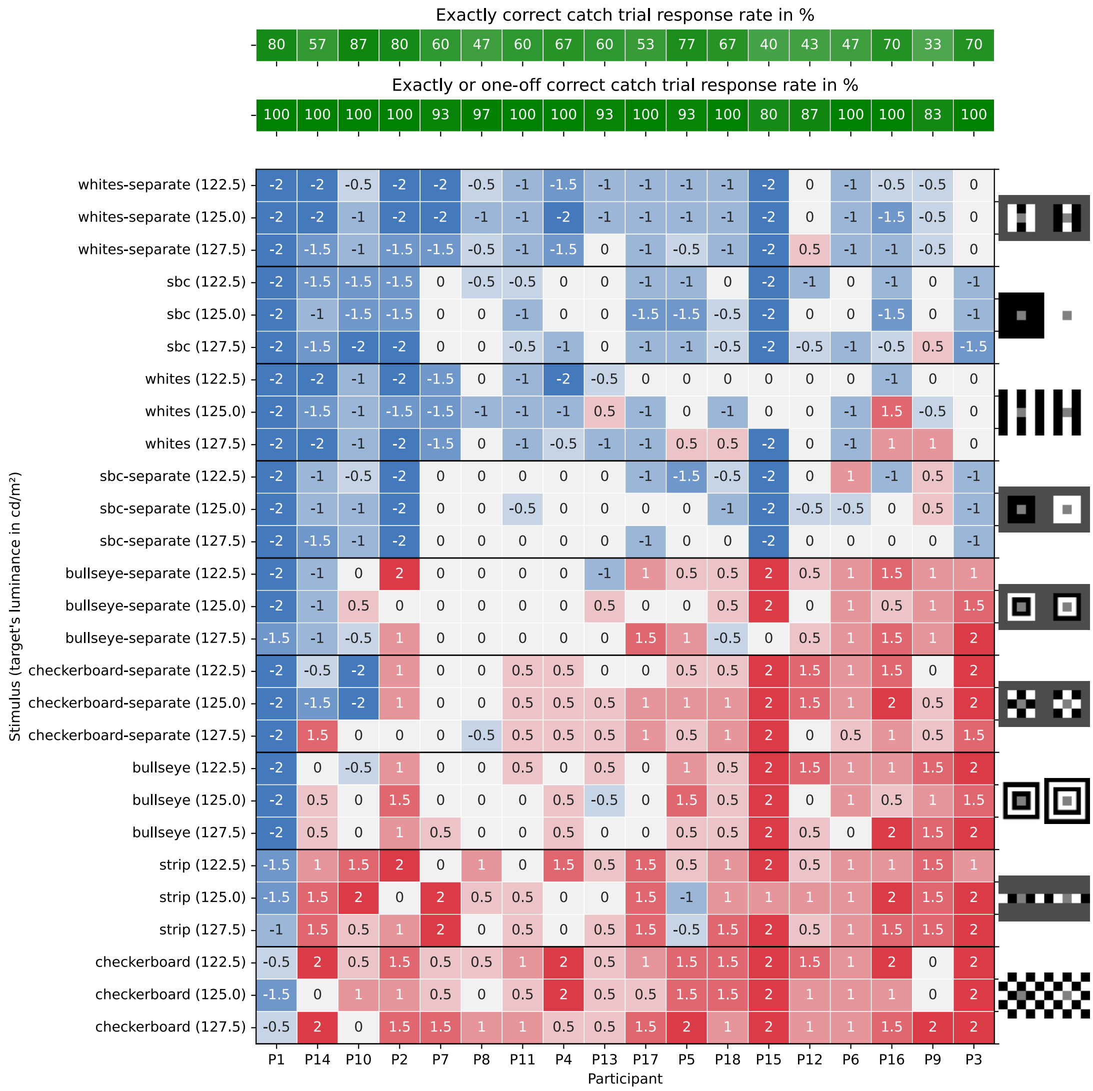


Figure 5.7: Brightness Ratings: Heat map representing the median response of participants to varied stimuli at different luminance intensities. The y-axis delineates the stimuli, arranged in ascending order based on their average response. The x-axis at the bottom lists the participants, arranged in ascending order based on their average response across stimuli. On the top there are two additional x-axes. The exactly correct catch trial response rate is the rate where participants responded with the exact expected response to the catch trial. Below that, The exactly or one-off correct catch trial response rate is the rate where participants responded with the exact expected response or +1 or -1 of the exact expected response.

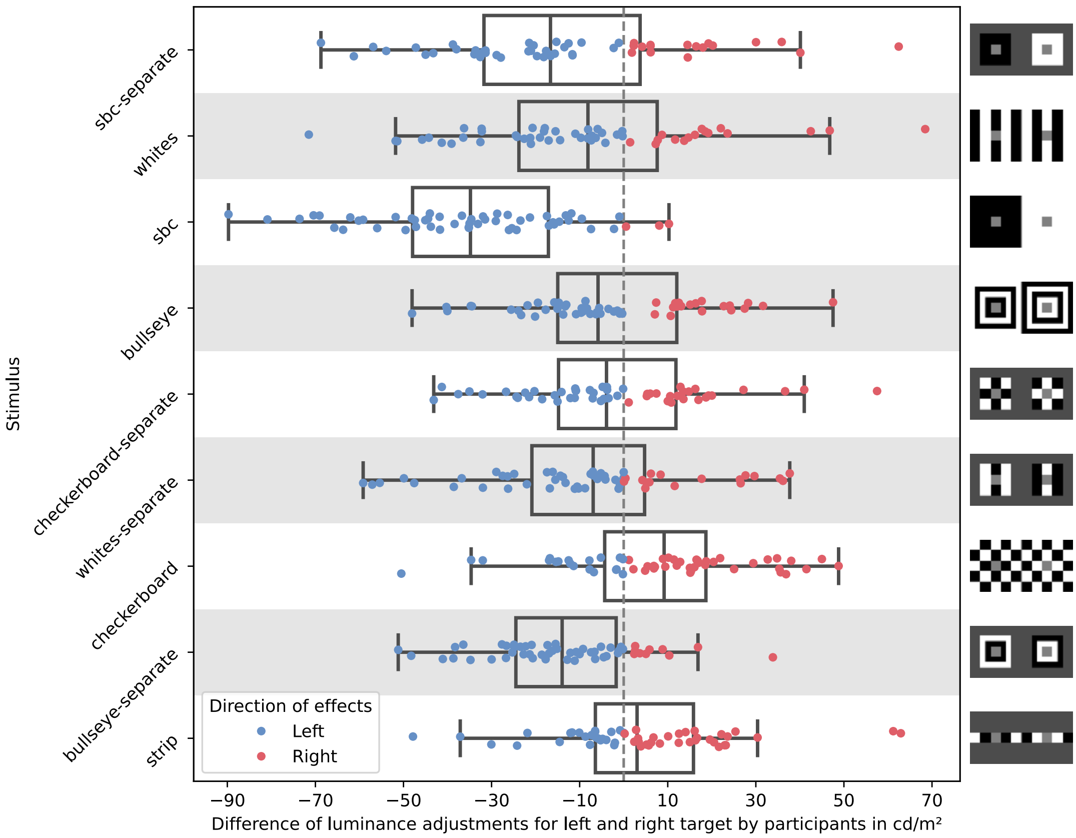
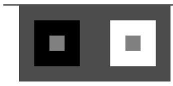
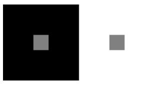
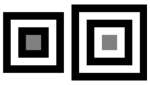
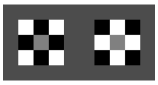
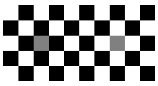
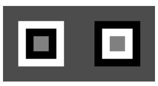


Figure 5.8: Method of adjustment: Differences of the luminance adjustments made by participants to match stimulus targets. The data points represent adjustment differences for left and right targets. Blue represent the negative adjustment differences and red represent the positive adjustment differences. Stimulus arranged on the y-axis, are ordered based on the interquartile range. The x-axis shows luminance in cd/m². If data points are positioned exactly on the the vertical dashed line positioned on 0 cd/m², it means both left and right target were adjusted with the same luminance. Alternating gray and white horizontal bars on the background are for guidance purposes to easily distinguish the data points of one stimulus.

5.3.3 Comparing Variability of Brightness Effects

While the method of adjustment offers a wide range of variability, brightness ratings largely cluster most stimuli around similar values. For most stimuli in the method of adjustment, there is an increasing variability trend, while in brightness ratings, interquartile range values remain relatively static at 1.

Table 5.3: The calculated interquartile range for the respective stimuli in both methods.

Stimulus		Interquartile Range	
		Brightness ratings	Method of adjustment (in cd/m²)
	“sbc-separate”	1	35.45
	“whites”	1.25	31.40
	“sbc”	2	30.84
	“bullseye”	1	27.03
	“checkerboard-separate”	1	26.65
	“whites-separate”	1	25.59
	“checkerboard”	1	22.98
	“bullseye-separate”	1	22.78
	”strip”	2	22.27

Chapter 6

Discussion

The main goal is to find whether the results measured with two different methods, the method of adjustment and brightness ratings, give similar estimates. This can help researchers understand the methods better when selecting one for their specific studies. The limitations when comparing the data collected in prior studies are that the similarity of the results could be due to the variations of the stimuli or the individuality of the participants. Because prior studies used brightness illusions in different variations and involved different participants, conducting the experiment using both methods with the same brightness illusion and the same participant resolved previously identified limitations.

Overall, the analysis reveals a similar average direction of effect for the majority of the brightness illusions, this includes three from the four original stimuli and two from the five novel stimuli, the analysis shows a similar average direction of effect. The reason not all brightness illusions have the same average direction of effect is that the method of adjustment provides results that differ from what is expected for three stimuli: "bullseye", "bullseye-separate," and "checkerboard-separate". Variability among the participants could influence this result. On the other hand, there is minimal similarity in the strength and variability of effects for all stimuli.

For stimuli that are highly similar to those from prior studies, there is complete agreement on the average direction of effect, as shown in Table 6.1 and Table 6.2.

Table 6.1: Comparison of average direction of effect for similar stimuli in present and prior work using method of adjustment. A checkmark (✓) indicates agreement in the direction of effect between the present and prior work for the given stimuli when the surrounding context is on the same side.

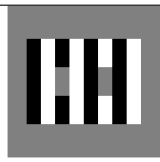
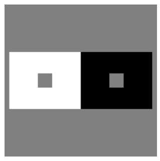
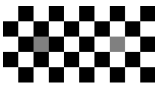
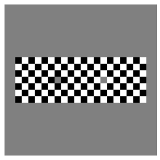
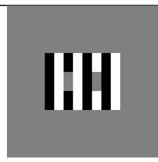
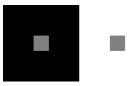
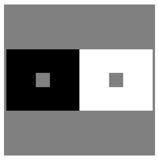
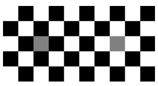
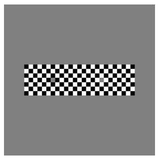
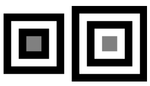
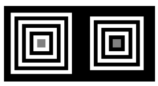
Method of Adjustment			
Present work			Prior work
In agreement	Stimulus from present work	Average direction	Stimulus from prior work
✓	 “whites”	Left Left	 ”White-groß”
✓	 “sbc”	Left Right	 ”Simultaner Helligkeitskontras”
✓	 “checkerboard”	Right Right	 ”Schachbrett-groß”

Table 6.2: Comparison of average direction of effect for similar stimuli in present and prior work using brightness ratings. A checkmark (✓) indicates agreement in the direction of effect between the present and prior work for the given stimuli when the surrounding context is on the same side.

Brightness ratings				
Present work				Prior work
In agreement	Stimulus from present work	Average direction		Stimulus from prior work
✓	 “whites”	Left	Left	 ”r_WE_thick”
✓	 “sbc”	Left	Left	 ”r_sbc_large”
✓	 “checkerboard”	Right	Right	 ”r_checkerboard _0938”
✓	 “bullseye”	Right	Left	 ”d_bullseye”

6.1 Limitations

The present work has some limitations. Based on the collected data, there were some data that had issues originating from specific participants. Therefore, data from these specific participants were excluded from the analysis in both methods.

One possibility for the different results in the average direction of effect for the three stimuli, “bullseye”, “bullseye-separate”, and “checkerboard-separate”, relates to their presentation. Although the stimuli were identical in size, shape, luminance, and contrast, they differed in one key aspect determined by the methods used: the number of targets shown in each stimulus. In brightness ratings, both targets were shown, while in the method of adjustment, only one target at a time was shown due to the method itself. This difference could lead the three stimuli to have the same effect that the “sbc” has, in which the target is only compared to its only immediate surrounding rather than the surrounding pattern. By leaving out the other target and replacing it with either black or white, the focus of comparing elements can shift from pattern-oriented comparison, in which the target is compared to the surrounding pattern, to a combination of comparing the target with its immediate surrounding and just comparing the target with the black or white square that replaced the counter target. This can be supported by the observation that stimuli with targets encircled completely by only black or white always have the same average direction of effect in the method of adjustment. This can be a motivation for further studies where the data collected with the method of adjustment showing one target and showing both targets are being compared.

Also, the order of conducting both methods could influence this result. Because participants who started the experiment with brightness ratings already know how the stimuli would look with both targets, while participants who started with the method of adjustment first only see the stimuli with one target at a time. Upon checking the method order, based on the time stamps and their results, there is no striking evidence to suggest a relationship between the order of the methods and the average direction of effect.

6.2 Individual Differences

Participant P21 did not follow the instructions in both methods. In the method of adjustment, as shown in Figure A.6, this participant adjusted to match the opposite side of the only target shown. So the adjustments of this participant were either black or white. In the brightness ratings method, this participant has a very low exactly correct catch trial response rate of 23%, meaning only 23% of the catch trials were done correctly. Even if the responses are inverted, the rate does not change. Because of these reasons, participant P21 was excluded from the analysis in both methods.

There is another participant P18, as shown in Figure A.5, who even had a lower rate of exactly correct catch trial responses of 13%. However, if P18’s responses are inverted, the rate changes to 67%. Upon examining the individual values from each trial by P18, it further supports the suspicion of inversion, as almost all of the responses are the direct opposite of the average. This evidence strongly suggests that P18 inverted the task, prompting them to determine which target appeared darker. In the method of adjustment, Participant P18 does not have any particularly striking values compared to the average. For these reasons, the responses from Participant P18 in brightness ratings are inverted in the analysis.

Another candidate whose input is strange is participant P20. This participant also has a very low rate of exactly correct catch trial responses, at 17%, as shown in Figure A.5. However, unlike participant P18, the inverted responses of participant P20 still have a low rate of response of 43%. Additionally, the rate of exactly or one-off correct catch trial responses is the lowest. Due to these findings, participant P20 was also excluded from the analysis.

The last participant excluded was participant P19. In the method of adjustment, this participant's adjustments are very low compared to the average, as shown in Figure A.6, affecting the average adjustments from the entire experiment. Just like participant P21, participant P19 also had adjustments that were only black. Because of these significant and consistent outliers, data from P19 was also excluded from the analysis.

Another limitation is that participants may not use the entire range of options in brightness ratings, whereas in the method of adjustment, participants may not use the buttons to adjust the stimulus in small intervals as this takes more time. Given the time recordings, it could be an indication of cognitive laziness.

Moreover, three participants, P1, P2 and P3 are participants that have done this kind of brightness perception experiments before. This could lead to expectancy effects, where their previous experiences influence their perception and expectations in the current experiment, according Bar (2007). But as shown in Figure 5.7, from all participants, participant P1 seem to gave responses that are not aligned with the expected outcome.

6.3 Implications and Applications

The results address the research question and highlight the potential for further exploration. It may prove beneficial to examine whether a modified method of adjustment yields different results by displaying both targets of a stimulus and employing an alternative method to identify the reference target. This could have an impact on the final outcome of an experiment.

In addition, the majority of participants in the debriefing reported that the brightness rating task was much easier to perform than the method of adjustment task. Depending on the order of the method, this could influence a participant's accuracy in doing the tasks correctly in one method or the other.

The results also show that there is greater variability in the method of adjustment that was not captured in the brightness ratings. If the focus of a study is on individual variations, the method of adjustment might be the better option to choose as the method for the study. For a more robust method with respect to individual outliers, brightness ratings would be the better option. This is supported by the example that in this experiment, when the excluded data of three participants are included, the result collected with the method of adjustment changes for many stimuli, for example, affecting the average direction of the effect. However, the data collected with the brightness ratings change only minimally. In fact, only two stimuli change the average direction of effect results from no direction of effect to a rightward direction of effect.

Apart from this, the number of participants shows that the lower the number of participants, the more prone the overall result is to outliers.

6.4 Conclusion

In conclusion, this present work fills the missing appropriate comparison of the results from two distinct methods to measure human brightness perception, brightness ratings and the method of adjustment. For this purpose, both methods utilized the same brightness illusions and involved the same subset of human participants. Furthermore, this study confirms the similarity of the results from two prior studies regarding the average direction of effect from Allaham (2022) and Bindermann (2022) for the the majority of the original brightness illusions. Based on the inputs from some participants, it also shows that human brightness perception can be individualistic in both methods. The dissimilarity in the average direction of effect for some stimuli including the novel stimuli, suggests a need for further research. Nonetheless, the data collected with the method of adjustment and brightness ratings give estimates where a similarity can be recognized among the human participants, indicating humans might reflect the same perceptual behavior.

References

Adelson, E. H. (1995). *Checkershadow illusion*. Retrieved from http://persci.mit.edu/gallery/checkershadow. (Cit. on pp. 15, 16)

Adelson, E. H. (2000). *Lightness perception and lightness illusions*. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press. (Cit. on p. 15).

Allaham, M. A. (2022). *Investigating inter-individual differences in human brightness perception* (Bachelor's thesis, Technische Universität Berlin). (Cit. on pp. 17, 18, 21, 22, 24, 25, 29, 56).

Bar, M. (2007). *The proactive brain: Using analogies and associations to generate predictions*. *Trends in Cognitive Sciences*, *11*(7), 280–289. (Cit. on p. 55).

Bindermann, M. (2022). *Vergleich ausgewählter helligkeitsphänomene: Modellbasierte vorhersagen und psychophysische messungen [comparison of selected brightness phenomena: Model-based predictions and psychophysical measurements]* (Bachelor's thesis, Technische Universität Berlin). In German. (Cit. on pp. 17, 19, 21, 23–25, 29, 56).

Brunn, C. (2020). *The crispening effect: An artifact of a method or a feature of the visual system* (Bachelor's thesis, Technische Universität Berlin). (Cit. on p. 15).

Cunningham, D., & Wallraven, C. (2011). *Experimental design: From user studies to psychophysics*. CRC Press. (Cit. on p. 32).

Domijan, D. (2015). *A neurocomputational account of the role of contour facilitation in brightness perception*. *Frontiers in Human Neuroscience*, *9*, 93. (Cit. on pp. 17, 21, 23, 27, 28).

Fechner, G. (1860). *Elemente der psychophysik*. Leipzig: Breitkopf und Härtel. (Cit. on p. 17).

Goodwin, K. A., & Goodwin, C. J. (2016). *Research in psychology: Methods and design*. John Wiley & Sons. (Cit. on p. 35).

Murray, R. F. (2020). *A model of lightness perception guided by probabilistic assumptions about lighting and reflectance*. *Journal of Vision*, *20*(7), 28–28. (Cit. on p. 21).

Robinson, A. E., Hammon, P. S., & de Sa, V. R. (2007). *Explaining brightness illusions using spatial filtering and local response normalization*. *Vision Research*, *47*(12), 1631–1644. (Cit. on pp. 17, 21, 23, 27, 28).

Appendix

A.1 Median and Mean in Brightness Ratings

Figure A.1 shows the brightness ratings data with another alternative statistical measure which is the mean (arithmetic average). All stimuli in Figure A.1 have the same average direction, as shown in Figure 5.1, except for the stimulus “sbc-separate” and stimulus “bullseye-separate”. In these cases, the arithmetic mean suggests a direction effect, while the median does not. Both novel stimuli have the same average direction of effect as their original stimuli.

A.2 Catch Trial Responses in Method of Adjustment

Figure A.2 depicts the arithmetic mean of luminance adjustments for the targets in catch trials of the adjustment method experiment. Notably, participants P1, P2, and P3 demonstrate lower luminance adjustments compared to the overall average. Given that these participants had previous experience with brightness perception experiments, this pattern may suggest that experience with similar experiments could influence luminance adjustment behavior over time.

A.3 Adjustments to the Different Targets and Intensity Values in Method of Adjustment

Figure A.3 displays each adjustment to the different intensity values from participants as an individual data point. It shows differences in the adjustments to different intensity values.

Figure A.4 presents the mean luminance adjustments of each target in a stimulus. It shows that the distances from the actual luminance to the luminance adjustments for the left and right target are not the same. For most stimuli the luminance adjustments were lower adjusted than the actual luminance of the target.

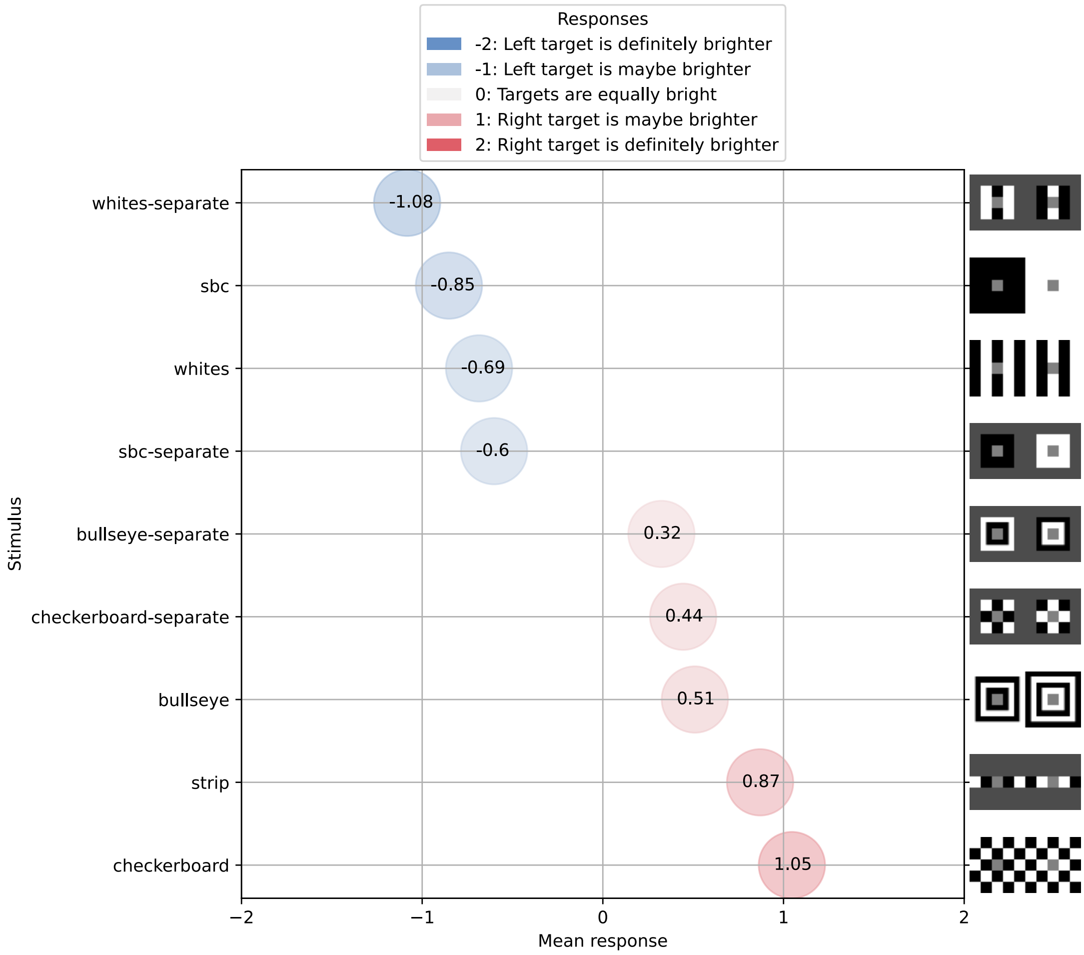


Figure A.1: Unfiltered data from brightness ratings: Each data point presents the mean response from all participants to the respective stimulus. The y-axis delineates the stimuli arranged in ascending order based on their average response. The x-axis at the bottom shows response values. Negative average responses are colored bluish and indicate a leftward direction of effect, while positive average responses are colored reddish and indicate a rightward direction of effect. The more saturated the colors, the more certain the response is.

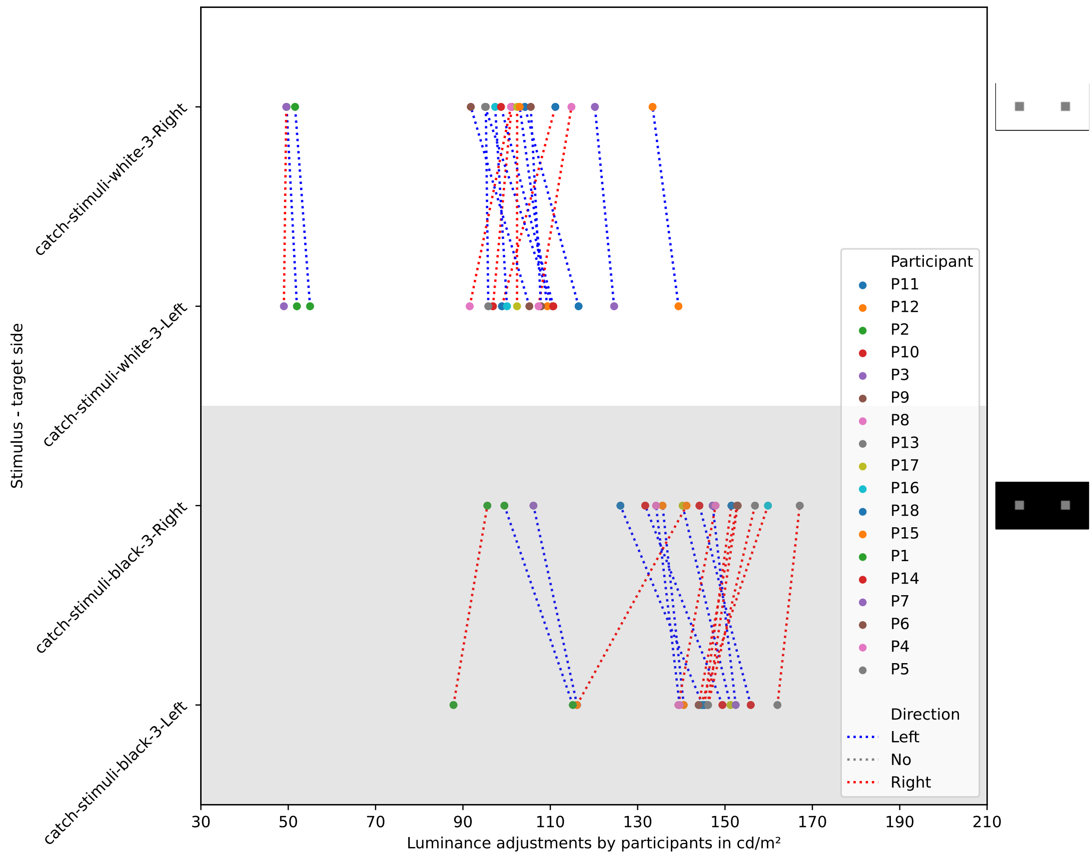


Figure A.2: Method of adjustment: Each pair of dots connected by a dashed line represents a participant's average adjustment for the right and left target. The y-axis shows catch stimuli. The x-axis displays luminance in cd/m².

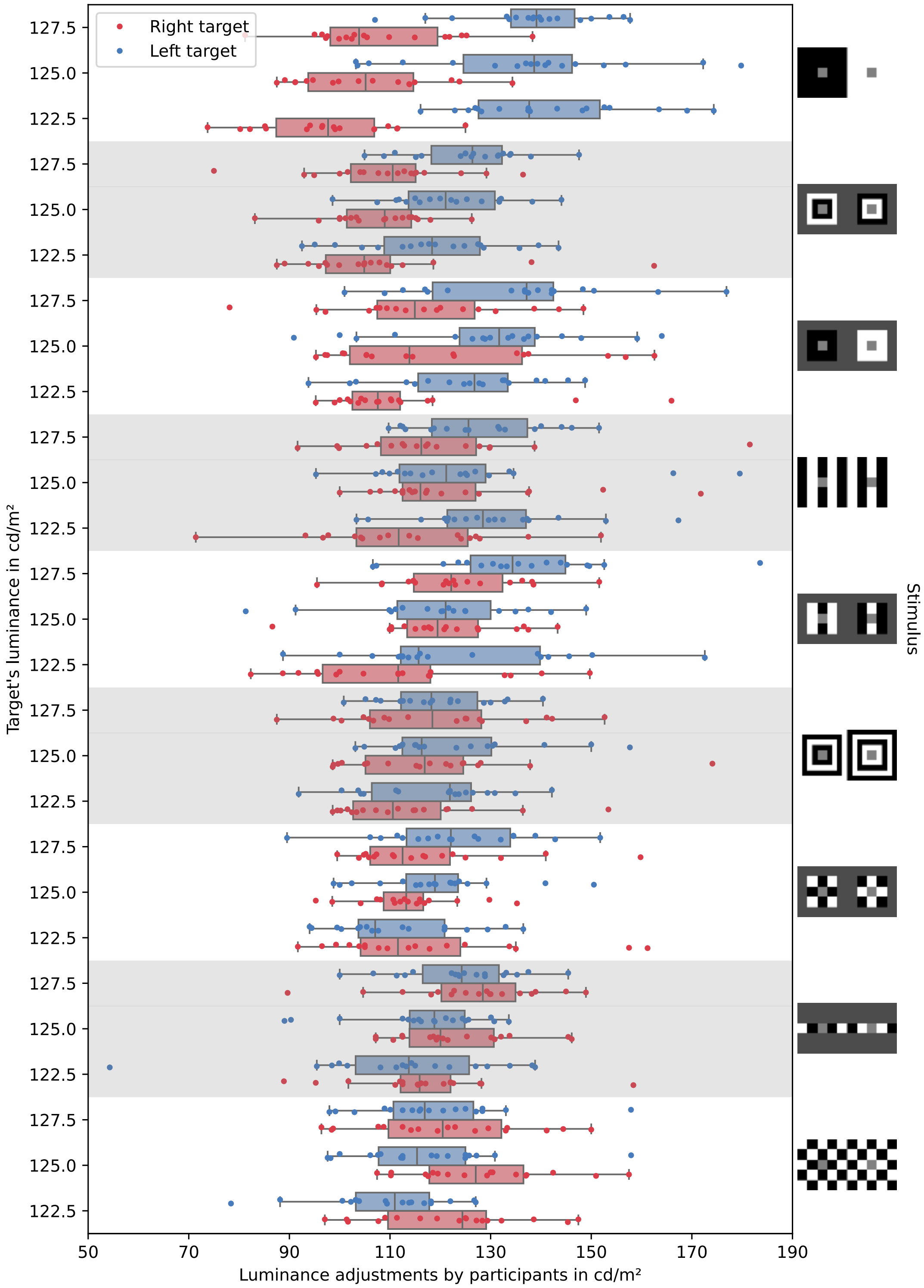


Figure A.3: Method of adjustment: Data points represent luminance adjustments participants made to match stimulus targets. Blue and red represent adjustments for left and right targets, respectively. Stimuli, arranged on the y-axis, are based on the difference in luminance adjustments between the targets, with each having luminances of 122.5, 125, and 127.5 cd/m². The x-axis shows luminance in cd/m².

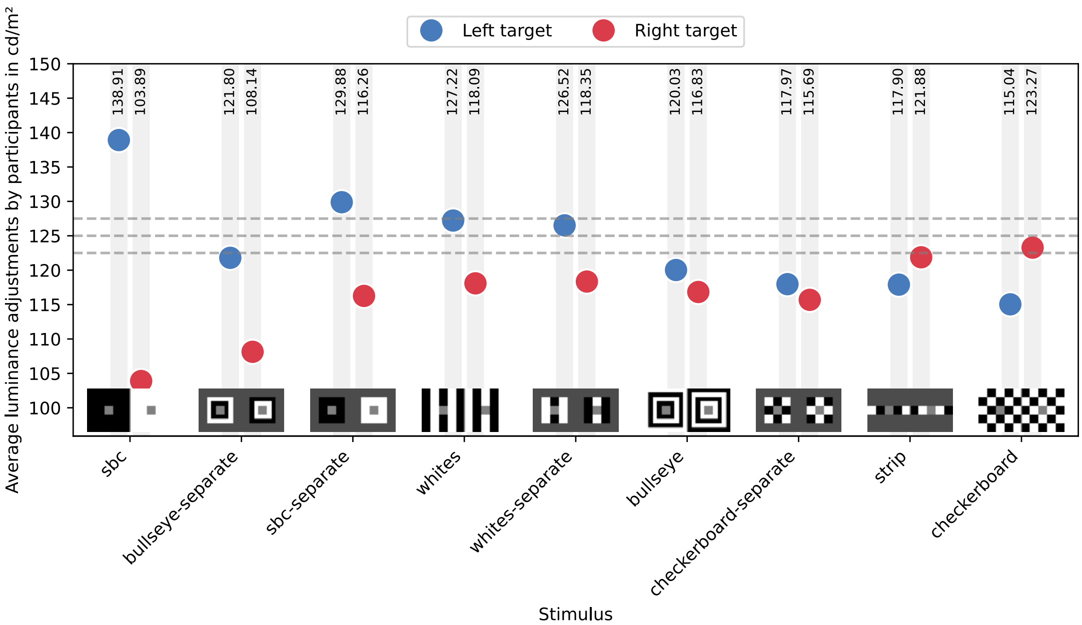


Figure A.4: Method of adjustment: Average (mean) adjusted luminance to match the corresponding target. The x-axis lists stimuli with their left and right target in ascending order, based on the difference of the luminance adjustment between the targets. The y-axis scales the luminance in cd/m². Blue data points represent the average luminance adjusted to match the left target, while red dots signify the same for the right target. The dashed horizontal lines represent the actual luminances of the target that were presented to the participants during the experiment trial.

A.4 Unfiltered Data plots

The following figures in this section show unfiltered data plots as referenced in the discussion chapter. As mentioned earlier, data from three participants was discarded from the analysis and the responses from one participant were inverted. In this section, the data plots are shown unfiltered which means the data from the three participants P19, P20, and P21 are included, and the data from P18 is not inverted.

Upon examination of Figure A.5, which displays the exact catch trial response rates, one observes that Participant P19, possessing a rate of 40%, represents the highest among those excluded. However, participant P9 has an even a lower rate with a 33%. The decision to exclude Participant P19 is based on findings from the method of adjustment. To maintain consistency, when data from one method for a participant is excluded, data from the other method from the same participant is also excluded. This protocol ensures uniform treatment of data across the study.

The findings of the data form participants P19 and P21 within the method of adjustment task present anomalies, as seen in Figure A.6 and Figure A.7.

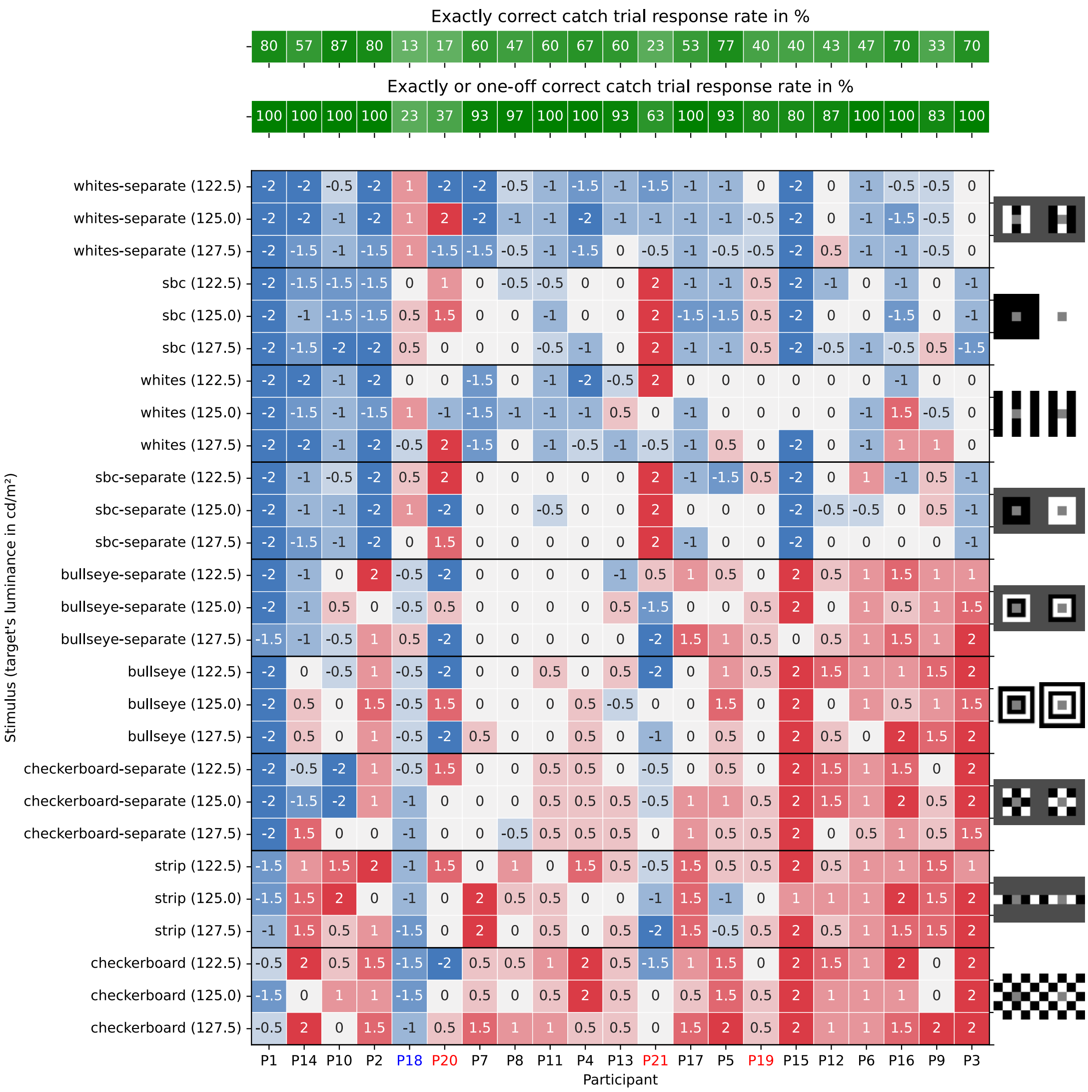


Figure A.5: Unfiltered data from brightness ratings: Heat map representing the median response of participants to varied stimuli at different luminance intensities. The y-axis delineates the stimuli, arranged in ascending order based on their average response. The x-axis at the bottom lists all participants, arranged in ascending order based on their average response across stimuli. Red-marked participants were discarded, and blue-marked ones had inverted responses in the analysis. On the top there are two additional x-axes. The exactly correct catch trial response rate is the rate where participants responded with the exact expected response to the catch trial. Below that, The exactly or one-off correct catch trial response rate is the rate where participants responded with the exact expected response or +1 or -1 of the exact expected response.

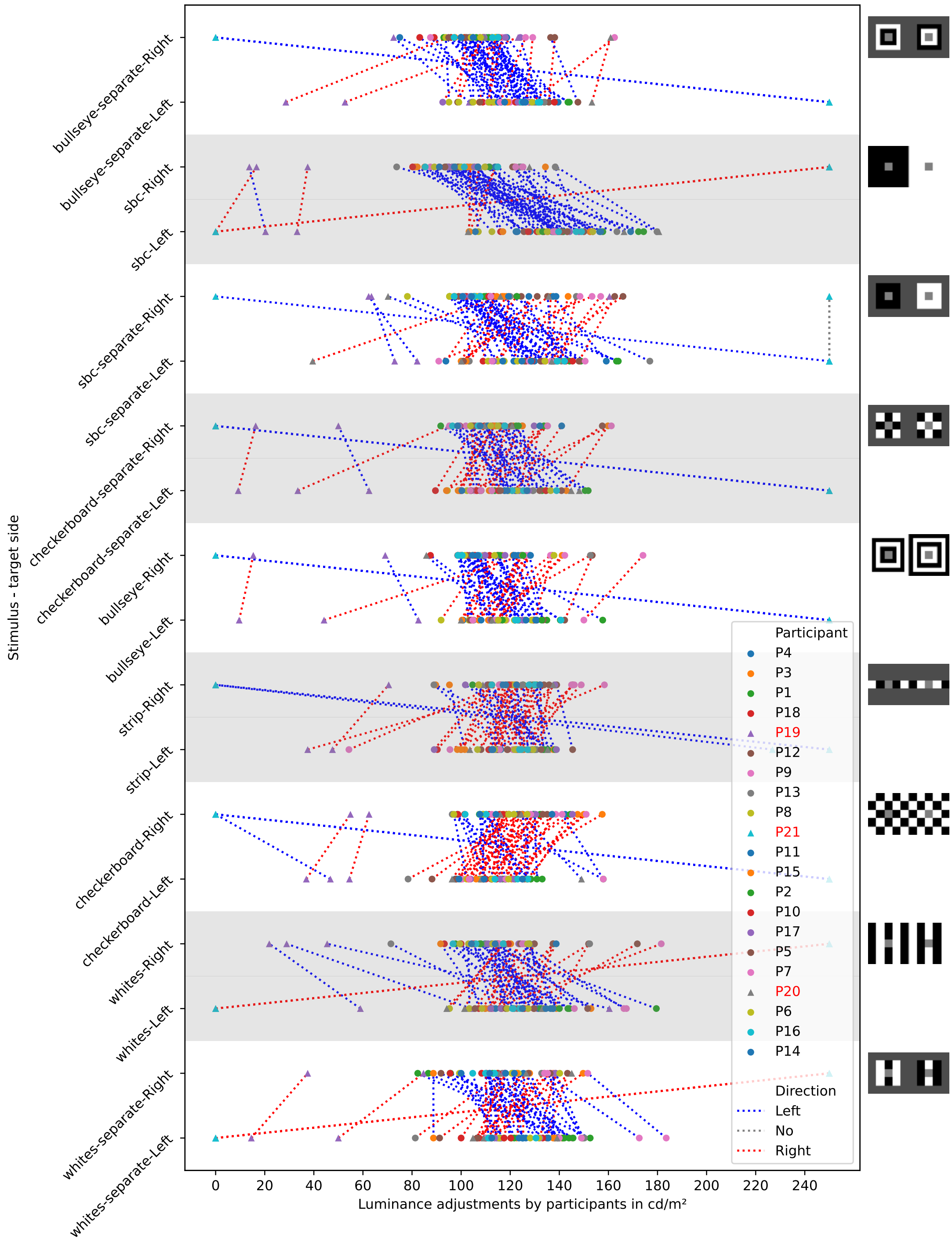


Figure A.6: Unfiltered data from method of adjustment: Each pair of dots connected by a dashed line represents a participant's adjustment for the right and left target. The y-axis shows stimuli in ascending order based on luminance difference between left and right targets. The x-axis displays luminance in cd/m².

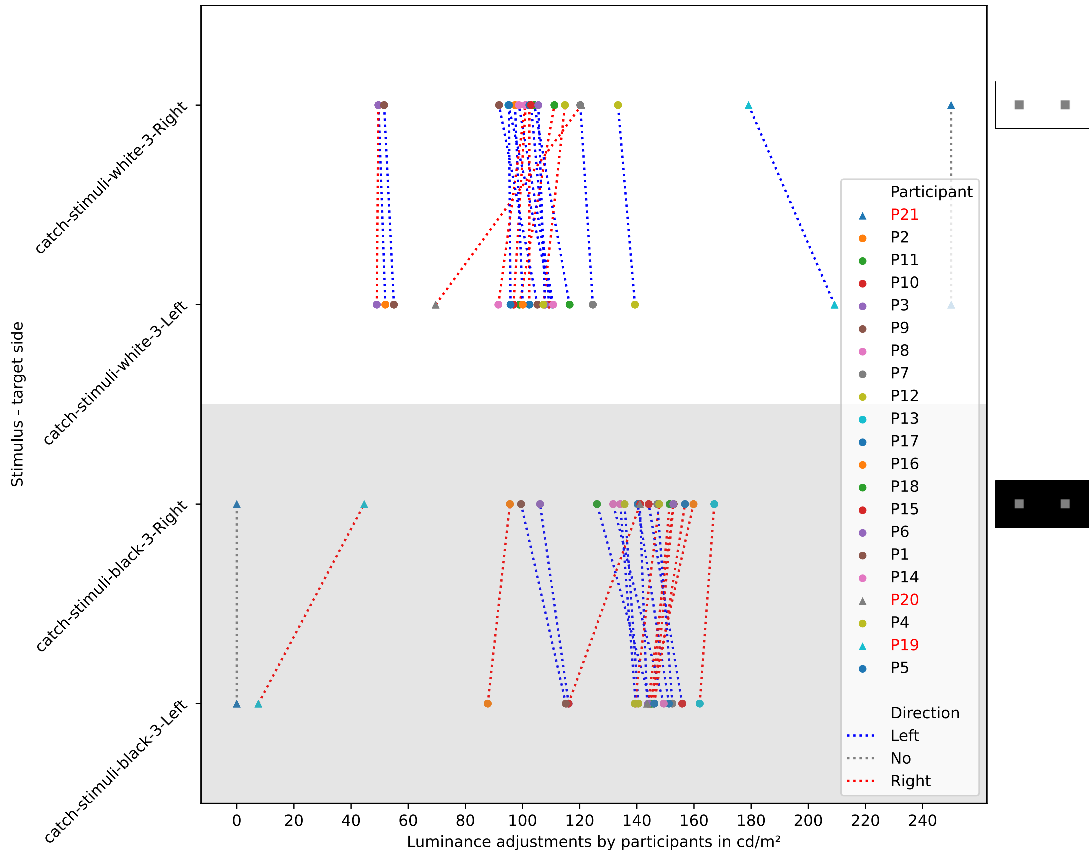


Figure A.7: Unfiltered data from method of adjustment: Each pair of dots connected by a dashed line represents a participant's average adjustment for the right and left target. The y-axis shows catch stimuli. The x-axis displays luminance in cd/m².

A.5 Statement of Ethics Committee

ethikkommission@eecs.tu-berlin.de
Ethik-Kommission (EK) der Fakultät IV
Technische Universität Berlin
Sek. MAR 5-5
Marchstr. 23, 10587 Berlin

Berlin, 09.08.2023

Application ID: 254
Submitted at: 02.08.2023
Business reference: -

Scientist in charge:

Prof. Dr. Marianne Maertens

Title/description of the research project/study:

"Investigating Different Methods to Study Human Brightness Perception"

Statement of Ethics Committee for self disclosure in standardized questionnaire

The aims and procedures of this study – as reported by the scientist(s) in charge in a standardized questionnaire – do not require a statement of the ethics committee.

Research participants

- are not expected to take emotional or physical risks,
- are not exposed to a high level of (physical or emotional) stress,
- are fully informed about the purpose and procedures of the study,
- do not belong to a vulnerable group (e.g. children, patients, persons with limited capability to judge)
- do not undergo fMRI (functional Magnetic Resonance Imaging) or TMS (Transcranial Magnetic Stimulation) during the course of the study

The scientist(s) in charge have agreed to protect data privacy, to fully inform participants about the purpose and procedures of the study, and to obtain informed consent prior to the study. This statement presupposes the truthful reply to all questions in the questionnaire and the responsible consideration of all ethical aspects of experimental study design with human participants.

Kind regards,

Ethics Commission Faculty IV
TU Berlin

1 / 1

A.6 General Information for Participants

Technische Universität Berlin
Computational Psychology
Sekretariat MAR 5-5
Marchstr. 23, 10587 Berlin
Tel. +49 30 314-75771

Contact person for any queries:
Hafidz Arifin (h.arifin@campus.tu-berlin.de)

General Information for Participants, Declaration of Consent and Statement on the Storage of Personal Data

Thank you for your willingness to participate in the study 'Investigating Different Methods to Study Human Brightness Perception'. This experiment is about measuring human brightness perception and how humans perceive brightness.

Procedure of the Study
Participants (subjects) of this experiment sit in a darkened room in front of a computer monitor. The subject is shown a total of 150 test images. These test images consist of black and white areas and gray fields. The subject's task is to follow the instructions according to the method. There are no right or wrong answers.

> <u>Instructions for Method 1: Brightness Matching</u>
> Adjust the brightness of the square target field above the test image so that it matches the brightness of the indicated gray field in the image. Subjects can adjust the brightness of the square target field with a button controller and confirm their adjustment by pressing the middle button.
>
> <u>Instructions for Method 2: Brightness Rating</u>
> Rate which gray field is brighter in the test image. Participants can select an option with a button controller and confirm the selection by pressing the middle button.

After pressing the middle button the subject is shown the next image. Subjects will have the opportunity to take a break in between. In total, the experiment will last about an hour.

Voluntary and Anonymous
Participation in the study is voluntary. You may terminate your participation at any time without giving any reason and without any negative consequences. The data and personal communications collected as part of this study and described above will be treated confidentially. Those project staff members who have personal data through the direct contact with you are subject to a duty of confidentiality. The results of the study will be published anonymously, i.e. your data cannot be attributed to you personally.

1/2

Data Protection
I agree that my anonymized data can be stored and used for research purposes and published. The collection of your personal data is completely anonymous and treated confidentially, i.e. your name is not requested at any point. Your answers and results are stored under a personal code word. The anonymized data is stored for at least 10 years. The disclosure of personal data to third parties is excluded. You can request the deletion of the personal data we have collected from you at any time. To do so, I contact the contact person provided above.

Compensation
For participating in the examination you will receive an expense allowance in the amount of 12,50 € per hour. The compensation will be paid to you in cash. If you receive the compensation in cash, you must sign a receipt stating your name and address. All related information will be kept completely separate from the examination data.

I am aware that my participation primarily serves scientific purposes and may not bring me any direct personal benefit. I have been informed that my participation will only take place if I have signed this consent form. I was given sufficient time to do this and decide whether or not to participate in this study. I agree that the data collected from the experiment will be stored anonymously and evaluated for scientific purposes (cf. Art. 89 GDPR).

I hereby agree that my contact details will be stored in an internal file system in compliance with data protection policy. The data is stored exclusively to inform me about further follow-up studies that are being carried out. My consent is voluntary and I can revoke my consent at any time verbally or in writing without giving reasons. A revocation will result in the complete deletion of all personal data. In addition, upon request, I will receive information at any time about which personal data about me is stored in the internal file system and for what purpose.

I have read and understood the information above in full. I can get a copy upon request. I also had the opportunity to ask questions and they were answered to my satisfaction.

____
First and last name of participant

____
Address of participant

____
Place, date and signature of participant

____
Place, date and signature of study coordinator

2/2