The relationship of lightness illusions uncovered by individual differences and its advantage in model evaluation

Yuki Kobayashi[1, 2] & Arthur G. Shapiro[1]
1. American University, 2. Ritsumeikan University

Abstract

Computational models that explain lightness/brightness illusions have been proposed. These models have been assessed using a simplistic criterion: the number of illusions each model can correctly predict from the test set. This simple method of evaluation assumes that each illusion in the set is independent; however, since lightness illusions are not independent of each other, an unbalanced test set may distort the evaluation of models. Moreover, evaluating models with a single value obscures where the model's strength and weakness lies. We collected illusion magnitudes of various lightness illusions through two online experiments and investigated their correlations to identify underlying factors in these illusions. Experiment 1 identified three common factors reflecting assimilation, contrast, and White's effect. Experiment 2, with a different illusion set, identified two factors reflecting assimilation and contrast, and detected high independence of an illusion caused by remote luminance gradients. We then examined three well-known models using the outcomes of the experiments. This model test showed that the redundancy in the illusion sets does not markedly skew the evaluation of the models, but the predictions by some models have a strong bias towards the contrast effect. This study clarified that correlations of illusion magnitudes provide valuable insights into both illusions and models, and highlighted the need to assess models based on processes underlying lightness perception, rather than focusing on individual illusions.

Introduction

The perception of lightness stands as a particularly fascinating aspect of human visual processing. The perceived lightness of an object is not solely determined by the light intensity received by the eyes from that object; it is also greatly influenced by the luminance pattern of the broader surrounding. This influence may be contrastive, where the object's lightness appearance stands out more from its immediate surroundings (Rudd, 2010; Wallach, 1948; Yund & Armington, 1975), or assimilative, where it blends more with them (Agostini & Galmonte, 2002a; Bressan, 2001; Gilchrist & Annan, 2002). Moreover, lightness is also likely to be affected by remote regions that do not share boundaries with the object in the retinal image (Adelson, 2000; Agostini & Galmonte, 2002b). Researchers have reported numerous lightness illusions, illustrating the complexity of human lightness perception.

A complete theory of human lightness perception is expected to be able to rigorously explain all of the known lightness illusions. Although none of the existing theories have achieved this feat, one of the major approaches to the complete theory is to model lightness processing with computational image processing (Blakeslee & McCourt, 1999; Land & McCann, 1971; Robinson et al., 2007). While theories largely described in natural languages have also made a large contribution to this research field (Adelson, 2000; Bressan, 2006; Gilchrist, 2006; Gilchrist et al., 1999), computational models have been receiving broad favor because they are inherently rigorous and able to take any images as inputs (Blakeslee & McCourt, 1999; Land & McCann, 1971; Robinson et al., 2007). This nature enables clarification of whether computational models' outputs align with human perception and facilitates their comparison. (Kim et al., 2018; Kobayashi & Kitaoka, 2022; Murray, 2020; Robinson et al., 2007; Zeman et al., 2015). Evaluating and refining existing models is a crucial process to lead to the ideal theory, and it is more feasible with computational models' rigorousness.

In many cases, lightness/brightness computational models have been evaluated with the number of illusions correctly predicted by them (Kobayashi & Kitaoka, 2022; Murray, 2020; Nedimović et al., 2022; Robinson et al., 2007). This naïve counting is a simplistic benchmark because it assumes that each illusion is independent and of equal importance. However, there is no guarantee of their independence. Rather, some visual phenomena may be driven by the same underlying mechanism (Adams, 2007) and this potential commonality might distort the model evaluation by the naïve counting. For example, even if a model can predict five illusions, if four of those are driven by the same mechanism, then the model essentially accounts for only two mechanisms. If another model can explain three independent illusions, it should be considered superior to the first one. Moreover, evaluating models with a single value may obscure where the model's expertise and weakness lie. If we are able to identify classes of "similar" illusions, we can then determine which classes a model excels or falls short in explaining. The validity of the naive counting needs to be examined with more comprehensive understanding regarding the similarities and differences among lightness illusions.

The present study aimed to clarify the relationships among known lightness illusions using correlations of illusion magnitudes. There have been discussions about the theoretical classification of lightness illusions (Agostini et al., 2020; Gilchrist, 2006) but no attempts have been made to classify them with empirical data. Similar illusions that share an underlying process should exhibit consistent tendencies in individual differences; that is, an individual who experiences a strong effect in one illusion is likely to perceive a correspondingly strong effect in another similar illusion. Correlations of responses to visual tasks have been used for examining the similarity between different visual processes (Adams, 2007; Kobayashi et al., 2021a; Shaqiri et al., 2019) and for identifying common underlying mechanisms of them (Kaneko et al., 2018; Mollon et al., 2017; Peterzell, 2016; Peterzell et al., 2017, 1995; Takahashi et al., 2023). Measuring correlations of illusion magnitudes and classifying lightness illusions will enable us 1) to examine redundancy in an illusion set and 2) to give more detailed evaluations of models with information about which class of "similar" illusions are explained well. Moreover, organizing the relationships among illusions can 3) offer insights into

establishing a standard set of illusions commonly used for model evaluations. This is something several researchers have emphasized as necessary (Murray, 2021; Schmittwilken et al., 2023). A series of studies have found that correlations between different illusions are weak while those between variants of the same illusion are robust (Cretenoud, Francis, et al., 2020; Cretenoud, Grzeczkowski, et al., 2020; Cretenoud et al., 2019; Grzeczkowski, Clarke, et al., 2017). These studies employed mostly geometrical illusions that involve size, length, angle, or position perception. Our focus was exclusively on lightness and this increases the likelihood of identifying common factors among illusions that were thought to be "different". We used open data provided by a previous study (Nedimović et al., 2022) and conducted two online experiments with larger sample sizes. We found that known illusion sets can be classified into a few groups.

Analysis of open data

Nedimović et al. (2022) measured the magnitudes of lightness illusions in 85 participants using an online experiment. The authors employed eleven lightness illusions and two Mondrian images (see Figure 2 in the original paper), and they compared the data of the human observers with the outputs from several computational models. This open dataset was useful in confirming that the correlations of illusion magnitudes are sufficiently informative to reveal their relationships.

We directed our analysis toward the eleven illusory images (Figure 2 A-K in Nedimović et al., 2022), motivated by our interest in the correlations between illusion magnitudes. These images included four instances of simultaneous lightness contrast (SLC), three variants of White's effect (Clifford & Spehar, 2003; White, 1979, 1981), and four illusions where the targets appear to have a luminance closer to that of their immediate surroundings (Bindman & Chubb, 2004; DeValois & DeValois, 1990; Economou et al., 2015; Gilchrist & Annan, 2002). The authors posited that two of the four illusions (Reversed Contrast and Dungeon Illusion) are driven by perceptual grouping, similar to White's effect, and the remaining two (Checkerboard Illusion and Bullseye Illusion) predominantly rely on assimilation[1]. Therefore, they suggested that the eleven illusions could be composed of three primary effects: 1) the contrast effect, where the appearance of a target is distorted away from its immediate surroundings; 2) the reverse contrast effect, where the appearance of a target is distorted away from its perceptually grouped and remotely located surroundings; and 3) the assimilation effect, where the appearance of a target is distorted toward its immediate surroundings. This classification is grounded in existing literature and theoretical. While the authors primarily analyzed the averages of illusion magnitudes, they did not explore the correlations and classifications using a quantitative approach.

[1] Gilchrist (2006) argued that Checkerboard and Bullseye illusions can also be interptered as outcomes of perceptual grouping.

Methods and Results

The dataset from Nedimović et al. (2022) includes the perceived lightness (0-1 scale) of two target areas in each of eleven images, which was collected from 85 observers through the adjustment task conducted online. It was composed of 1,870 data points (11 images * 2 targets * 85 participants). Differences in the perceived lightnesses of the two targets define the magnitudes of the illusions. The authors reported that all eleven illusions exhibited significant effects in the expected directions.

We analyzed the correlations of the illusion magnitudes. Pairs of illusions that seem phenomenologically similar exhibited significant correlations (Figure 1). We further conducted an exploratory factor analysis (EFA). Because the magnitudes of each illusion were supposed to conform to a normal distribution, we employed maximum likelihood estimation as the factoring method. Very Simple Structure (VSS) and Minimum Average Partial (MAP) criteria suggested a two-factor model. Additionally, a parallel analysis using 95th percentile points of simulated random data also supported a two-factor model. A varimax (orthogonal) rotation was used because the factor correlations identified with an oblimin (oblique) rotation were small ($r < .3$). The independence of the factors was beneficial for evaluating the lightness computational models, discussed later in the Model Test section. The factor loading matrix of EFA showed a relatively simple structure (Table 1) while some of the illusions showed small commonalities (e.g., WhiteLarge). Factor 1 is thought to reflect a process underlying the contrast effect. The four SLCs and WhiteLarge demonstrated high loadings on this factor. Although WhiteLarge was thought to be one of the reverse contrast phenomena, it was judged to be in the same class as normal contrast figures. This is understandable because the targets in WhiteLarge share long edges with the flanking bars, which supposedly exert a contrastive effect on the appearance of the targets (Taya et al., 1995; Todorović, 1997). Factor 2 exhibited high loadings from illusions with assimilative effects. Although Nedimović et al. (2022) argued that the illusions in this class should consist of the reverse contrast phenomenon and the assimilation phenomenon, the data-driven classification based on the correlations did not agree with this theoretical classification.

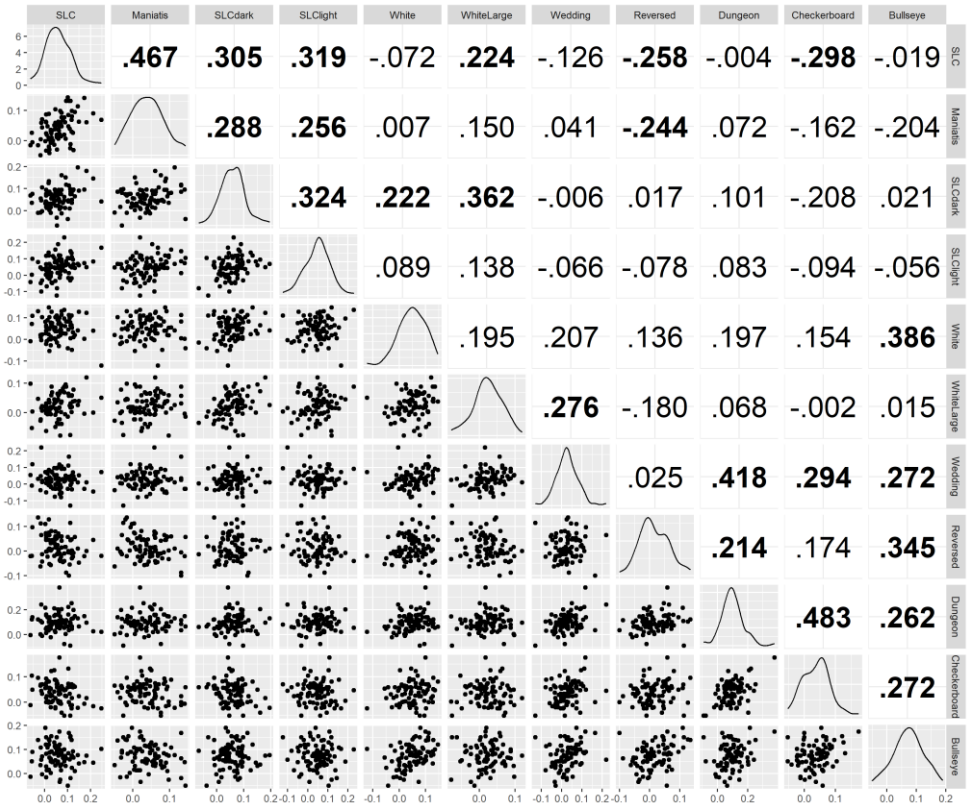

Figure 1. The correlation coefficients (upper triangle) and the scatter plots (lower triangle) of each pair of illusions. The coefficients in boldface indicate significant correlations (p .< 05; no corrections for multiple tests). See the original paper (Nedimović et al., 2022) for a detail of each illusion.

	Factor 1	Factor 2	Comm.
SLC	-.180	.665	.474
Maniatis	-.076	.600	.366
SLCdark	.061	.560	.317
SLClight	.002	.463	.215
White	.392	.121	.168
WhiteLarge	.163	.412	.196
Wedding	.549	.053	.304
Reversed	.303	-.257	.158
Dungeon	.701	.124	.506
Checkerboard	.619	-.262	.452
Bullseye	.481	-.071	.236

Table 1. The factor loading matrix of the data in Nedimović et al. (2022). The colors of the cells representing the factor loadings correspond to their values: orange for

positive and light blue for negative. The rightmost column indicates the commonalities (the sum of the squared factor loadings) of each illusion.

Discussions

The analysis of a public dataset (Nedimović et al., 2022) demonstrated that the individual differences in the magnitudes of lightness illusions are sufficiently informative for the classification of them. EFA indicated two classes, each composed of phenomenologically similar illusions. Factors 1 and 2 seem to correspond to assimilative and contrastive illusion groups, respectively. This result supports the effectiveness of EFA in classifying illusions. The fact that the four SLC variants consistently showed high loadings to Factor 1 agrees with previous studies that found robust intra-correlations of geometrical illusions (Cretenoud, Grzeczkowski, et al., 2020; Cretenoud et al., 2019; Grzeczkowski, Cretenoud, et al., 2017). While those studies also found that inter-illusion correlations are generally weak (Cretenoud et al., 2019; Grzeczkowski, Clarke, et al., 2017; Shaqiri et al., 2019), the current analysis, focusing specifically on lightness illusions, detected the similarity among assimilative illusions.

The factor structure indicated by the EFA may need consideration in the evaluation of lightness models. Nedimović et al. (2022) treated 11 illusions as eight items, but the classification by the EFA did not agree with this itemization. In the following two experiments, we examined larger image sets with larger samples, and performed model evaluations based on the results (Model Test section).

Experiment 1

Recognizing that online experiments and EFA can uncover an interpretable factor structure, we conducted our online experiment with a larger dataset on a commonly used set of illusions to examine their relationship. Here, we employed a set of illusions used by Robinson et al. (2007). They sought to extend an existing model, ODOG (Oriented Difference of Gaussian model; Blakeslee & McCourt, 1999), and proposed two new models: LODOG (Locally-normalized ODOG) and FLODOG (Frequency-specific Locally-normalized ODOG). Robinson et al. compared these models using the illusions that had been used to test ODOG (Blakeslee & McCourt, 2004; Howe, 2005; McCourt, 1982) and the number of successful predictions for them. This illusion set was also employed by a subsequent study (Zeman et al., 2015), with a substantial portion comprising variants of White's effect, suggesting redundancy. These model comparisons may cause a bias to White's effect and result in unfair evaluation of models.

Methods

We recruited participants via Prolific, an online recruiting service (https://www.prolific.co/), and compensated them with £3 for their participation. The participant pool of Prolific consists of individuals who are 18 years old or older and reside in

most OECD countries plus South Africa. We recruited 270 participants, ten times the number of items under study (Figure 2).

We created stimuli using Stimupy (Schmittwilken et al., 2023). Stimupy is a Python package to design stimuli for vision sciences, and it has a module to precisely recreate stimuli used by Robinson et al. (2007). Although Stimupy provides all illusions used in Robinson et al., we excluded two illusions because the targets were too small or narrow in these two figures to display in online environments. We ultimately used 27 stimuli (Figure 2). Each pixel was represented with a value ranging from zero to one. The targets and the gray surroundings of the illusions were represented with a value of 0.5 in the originally produced images, but because we converted the original images to correct gamma, the gray (0.5) was converted to 0.73 in the experiment. We assumed that the gamma value of monitors was 2.2 (raised to the power of 1/2.2). In most illusions, the other areas were represented with zero (black) and one (white), which remained unaffected by the current gamma correction. Corrugated-Mondrian (Adelson, 1993) and Grating-Induction (McCourt, 1982) were the exceptions. We employed an additional image (Figure 3A) as a filler stimulus, where two small square patches with different luminance levels were embedded in a black rectangle. The luminance levels of the two patches were approximately 0.67 and 0.78. This luminance difference was considered noticably large so the filler image was used to exclude participants who could not detect this difference.

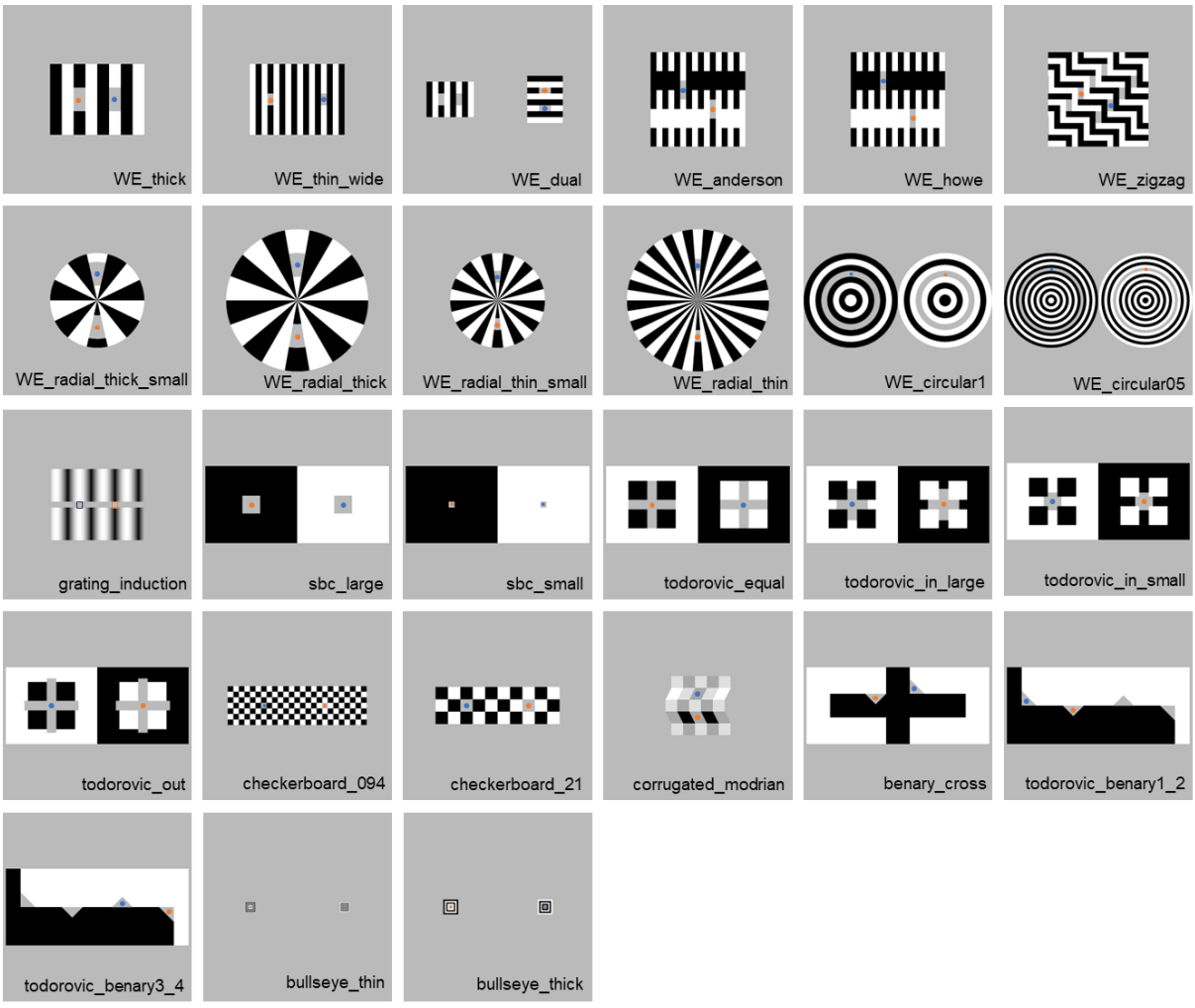

Figure 2. Twenty-seven illusions used in Experiment 1. Blue and orange dots (or frames) indicate target areas that were expected to be perceived darker and lighter, respectively (not shown in the experiment). Those in WE-howe, a control figure, were tentatively determined. We excluded WE-circular0.25 and Checkerboard-0.16 from Robinson et al. (2007) because the targets were too small or narrow in these two figures to display in online environments.

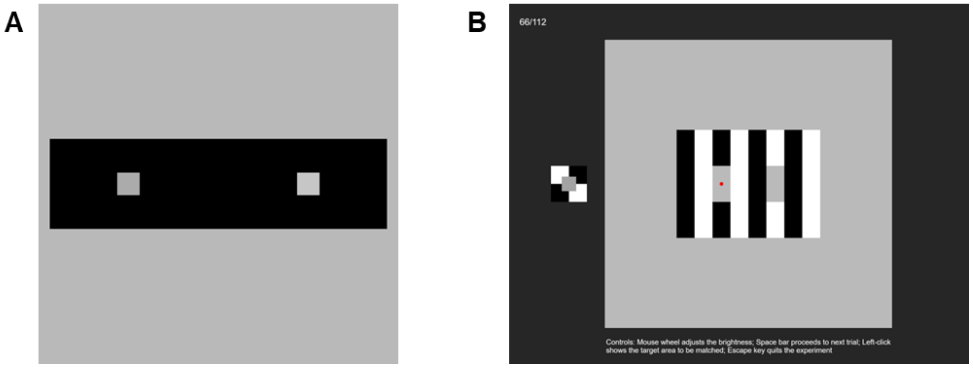

Figure 3. A) The filler image. The right small square was physically lighter than the left. B) An example of the display in the experiment. The count of trials (top left) and an instruction (bottom) were indicated throughout the experiment. The red dot on the figure indicated the target area.

The experiment was created with PsychoPy3 (Peirce et al., 2019) and was hosted on Pavlovia (https://pavlovia.org/), an online server for JavaScript-based programs. The participants accessed the experiment via their PCs; the use of tablets or phones were prohibited. A stimulus was displayed in the center of the screen in a square that occupied 80% of the screen's vertical length (Figure 3B). An adjustable patch on a white-black checker background was presented on the left of the stimulus. The center of the patch was located at a position shifted 50% of the screen's vertical length to the left from the monitor's center. The side lengths of the adjustable patch and the checkered background were 4% and 10% of the screen's vertical length, respectively. Initially, the patch was displayed with a luminance level randomly selected from a uniform distribution ranging from 0.58 to 0.85. The participants could manipulate its luminance with a mouse wheel in a range of 0.39 to 0.94. The amount of patch luminance change in response to the mouse wheel operation was also gamma-corrected to ensure a consistent amount of luminance change on the screen. The task was to adjust the luminance of the patch to that of a target area in the stimulus, indicated by a red dot at the beginning of a trial. The red dot appeared again when the participant left-clicked. When the participants were satisfied with their adjustments, they pressed a space key and started the next trial after a 0.5-second blank during which only dark background was presented. Each stimulus was presented either in the orientation shown in Figure 2 or rotated by 180 degrees. The entire session was composed of 112 trials (28 images × 2 rotations × 2 targets), whose order was randomized. The duration of the entire session was about 20 minutes. Before the session, the participants were instructed to "try to keep a constant distance from the screen throughout the

experiment." To motivate the participants to exert effort, we informed them that compensation might be withheld for “extremely inaccurate or random” task performance, although in reality, compensation was provided in all cases. The experimental program can be accessed via https://XXX, and the stimulus set is stored in the supplemental material in Open Science Framework (see Supplement).

Results

Differences of average adjustments for two targets were defined as the illusion magnitudes (orange minus blue in Figure 2). From 270 participants' data, we excluded 28 participants who produced illusion magnitudes falling outside the range of ±3 standard deviations for any of the 27 illusion items. Differences of adjustments for two squares in the filler image were also calculated (lighter minus darker) and 23 participants were excluded because they exhibited negative values. The remaining 219 participants were analyzed. These criteria for exclusion had been determined before conducting the experiment.

Table 2 shows the summary of the magnitudes. WE-Howe (Howe, 2005) did not show a significant illusory effect. That was expected because this figure served as a control. Since we were specifically interested in relationships of illusory effects and the sign of the magnitudes cannot be defined without a significant effect, we excluded WE-Howe from the following analysis. Most of the illusions produced effects consistent to our expectation but Todorovic-Equal showed an effect opposite to the report by Blakeslee and McCourt (1999). Its sign was reversed in the following analysis to ensure that positive values consistently represented effects aligned with human perception. The rightmost column indicates intraclass correlations ($ICC_{2,k}$), which examine the reliability of the measurements (Cretenoud et al., 2019). Each ICC was calculated using two magnitudes derived from two rotation conditions in each illusion. Therefore, the two measurements were not exact repetition but ICCs were generally high. A figure showing inter-illusion correlations can be found in the supplemental material.

	Mean	*SD*	*t*	*p*	*ICC*
WE_thick	15.95	18.79	12.56	<.001	.63
WE_thin_wide	21.85	20.44	15.81	<.001	.68
WE_dual	12.53	16.48	11.26	<.001	.49
WE_anderson	14.11	17.53	11.91	<.001	.62
WE_howe	0.78	19.06	0.60	.547	.66
WE_zigzag	11.81	18.31	9.55	<.001	.59
WE_radial_thick_small	7.68	17.57	6.47	<.001	.56
WE_radial_thick	11.01	18.25	8.92	<.001	.61
WE_radial_thin_small	16.93	19.49	12.85	<.001	.51
WE_radial_thin	18.04	21.00	12.71	<.001	.72
WE_circular1	32.21	21.73	21.93	<.001	.65
WE_circular05	40.85	25.14	24.04	<.001	.79
grating_induction	14.82	22.85	9.60	<.001	.68
sbc_large	8.14	21.34	5.65	<.001	.64
sbc_small	9.40	22.30	6.24	<.001	.57
todorovic_equal	-23.38	20.03	-17.27	<.001	.73
todorovic_in_large	19.06	18.68	15.10	<.001	.66
todorovic_in_small	13.75	15.49	13.14	<.001	.53
todorovic_out	21.26	20.73	15.18	<.001	.68
checkerboard_094	35.69	24.10	21.91	<.001	.73
checkerboard_21	24.74	17.66	20.73	<.001	.67
corrugated_mondrian	13.32	20.47	9.63	<.001	.66
benary_cross	6.04	16.73	5.35	<.001	.45
todorovic_benary1_2	7.05	18.36	5.68	<.001	.64
todorovic_benary3_4	3.94	16.27	3.58	<.001	.52
bullseye_thin	25.41	21.54	17.46	<.001	.72
bullseye_thick	36.37	25.21	21.36	<.001	.73
filler	-	-	-	-	.31

Table 2. Summary of the results. Mean and SD are based on magnitudes ranging from 0 to 255. ICC indicates $ICC_{2,k}$.

The Kaiser-Meyer-Olkin's measure of sampling adequacy was .77 for all the data and the minimum measure in each item was .51 in Todorović-Benary3-4. We interpreted this result as indicating that the sample deserved a factor analysis. The VSS and MAP criteria indicated seven and three factors, respectively. A parallel analysis using maximum likelihood estimation suggested three factors. Based on these criteria, we adopted a three-factor model. Table 3 shows the resultant factor loadings with varimax rotation. Some illusion pairs phenomenologically similar (e.g., SBC-large and SBC-small, WE-circular0.5 and WE-circular0.25, etc.) showed similar loadings and this suggests that the EFA could extract reasonable correlations. The three factors accounted for 27.8% of the total variance.

	Factor 1	Factor 2	Factor 3	Comm.
WE_thick	.195	.495	.123	.298
WE_thin_wide	.191	.571	.162	.389
WE_dual	.024	.225	.265	.122
WE_anderson	.224	.509	.180	.342
WE_zigzag	.306	.265	-.240	.221
WE_radial_thick_small	-.045	.311	-.075	.104
WE_radial_thick	.073	.440	.104	.210
WE_radial_thin_small	.068	.646	.014	.422
WE_radial_thin	.081	.623	.110	.406
WE_circular1	.633	.211	.005	.445
WE_circular05	.674	.215	.073	.505
grating_induction	.111	-.138	.326	.138
sbc_large	-.132	.022	.407	.183
sbc_small	-.224	.075	.545	.353
todorovic_equal	.503	-.072	-.169	.287
todorovic_in_large	.124	.237	.411	.241
todorovic_in_small	.096	.101	.402	.181
todorovic_out	.494	.091	.304	.345
checkerboard_094	.574	.318	.050	.434
checkerboard_21	.436	-.009	-.094	.199
corrugated_mondrian	.110	-.029	.298	.102
benary_cross	-.050	.250	.345	.184
todorovic_benary1_2	.026	.149	.522	.296
todorovic_benary3_4	-.185	.033	.216	.082
bullseye_thin	.478	.002	.305	.321
bullseye_thick	.588	.239	.103	.413

Table 3. The factor loading matrix obtained in Experiment 1. The cells are colored in the same manner as in Table 1.

Discussions

Experiment 1 confirmed that correlations of the illusion magnitudes can identify interpretable common factors from a set of lightness illusions. Apparently similar illusions exhibited similar patterns of factor loadings. Factor 1 was interpreted as a factor driving an assimilation effect. The illusions that showed their primary loadings on it were all assimilative. Factor 2 can be interpreted as related to T-junctions. The illusions primarily loaded onto this factor are plausibly driven by T-junctions and considered variants of White's effect (Howe, 2005; Yazdanbakhsh et al., 2002). Todorović-in-large, an illusion also considered as a variant of White's effect (Todorović, 1997), also showed a moderate loading on Factor 2. Factor 3 was considered as a factor for a contrast effect. While some illusions to which T-junctions appear relevant also demonstrated high loadings on Factor 3 (todorovic_benary1_2,

todorovic_in_large, todorovic_in_small), all four illusions explainable by simple contrast from immediate surroundings primarily loaded on Factor 3 (the two SBCs, Grating-induction, and Corrugated-Mondrian). The present experiment with a large sample size revealed that three simply interpretable factors underpin 26 lightness illusions.

The resultant classification offered some insights into lightness perception. First, as in the data of Nedimović et al. (2022), a factor for assimilation and that for contrast were extracted as independent factors. Some people may assume that a participant who perceives strong assimilation effects is likely to perceive weaker contrast effects, but these two illusion groups were not negatively correlated. They are not the other sides of the same phenomenon but are more likely to be independent processes. Second, it was suggested that White's effect has a distinctive characteristic. The substantial number of stimuli related to White's effect in this set likely increased the chances of identifying an independent factor for it, but it was intriguing that the empirical data indicated its uniqueness distinguished from assimilation or contrast. This factor seems to represent a process for T-junctions that modulates contrast effects from neighboring regions (Anderson, 1997; Blakeslee et al., 2016; Taya et al., 1995).

Moreover, the present result also suggested that illusions apparently rooted to the same phenomenon may actually be driven by different processes. For example, while Todorović-in-small, Todorović-in-large, todorović-equal, and todorović-out appear a continuum of similar illusions, two of them exhibited high loadings on the factor for contrast and the other two were dependent more on the factor for assimilation. This is consistent with a T-junction-based explanation proposed by the study that originally introduced these figures (Blakeslee & McCourt, 1999). Another example is circular versions of White's effect (WE-circular 05 and WE-circular 1). They were introduced to demonstrate that an illusion similar to White's effect arises even without T-junctions (Howe, 2005). However, the present experiment suggested that the illusory effect of these figures largely derives from a process different from the original White's effect, and again supported the idea that T-junctions characterizes it. The present analysis was particularly insightful in deepening the understanding of the cause behind apparently similar illusions.

It should be noted that the total variance of the three factors was not considerably large (27.8%). Even though the loadings on the three factors were significantly higher than those from random data, some illusions indicated the uniquenesses exceeding 80%. This partly hints independence of each illusion. However, it is also undeniable that there was variability in the level of uniqueness, suggesting the uneven importance as items for models to explain. More importantly, the explained variance depends on a sample size and the number of items. The larger sample size and the more items are used, the more likely low commonality is obtained (Gordillo et al., 2023). A further examination of the illusions' independence will be discussed in the Model Test section.

Experiment 2

We conducted another experiment using a different set of illusions. In this experiment, we included several illusions presumed to be important that were not part of the set used by

Robinson et al. (2007). Experiment 2 would provide insight into establishing a standard illusion set for model tests (Murray, 2021; Schmittwilken et al., 2023).

Methods

Experiment 2 was also conducted online using Prolific and Pavlovia. The sample size was determined to be ten times the number of illusions used. We aimed at 140 participants but obtained 141 people's data due to a technical reason. £2 was compensated for participation. There was no participant overlap between Experiment 1 and Experiment 2.

The experimental settings were almost the same as those of Experiment 1 except for the illusions set (Figure 4). First, we included two representative illusions from each of the three classes suggested in Experiment 1: WE-thick and WE-radial-thin-small from the class of White's effect; checkerboard 21 and WE-circular 1 from the class of assimilation; and SBC-large and Corrugated-Mondrian from the class of contrast. In addition, we employed illusions that are deemed as reverse contrast phenomena caused by perceptual grouping (Agostini & Galmonte, 2002a; Bressan, 2001; Economou et al., 2015; Gilchrist & Annan, 2002; Gilchrist et al., 1999). Reverse contrast was not identified as a factor independent of assimilation with the data by Nedimović et al. (2022), but we attempted to examine its independence again. For Agostni_cube, a simplified design by Domijan (2015) was used. In addition, an illusion introduced by Yazdanbakhsh et al. (2002) was employed as another example of White-effect variants. Similar to the circular version of White's effect (Howe, 2005), this illusion is also known as a counterexample of the T-junction-based explanation of White's effect (Yazdanbakhsh et al., 2002). We also included three illusions that contain explicitly "shadowed" regions: Koffka, simp-check, and snake (Adelson, 2000; Blakeslee & McCourt, 2012; Kobayashi & Kitaoka, 2022). They can be understood as variants of simlutaneous lightness contrast but the explicit shadow enhances an illusion in general. Lastly, Agostini_glare was employed as an illusion caused by luminance gradient (Agostini & Galmonte, 2002b; Kobayashi et al., 2021b; Zavagno et al., 2018). For the detail of each illusion, see the caption of Figure 4. The stimuli newly adopted in Experiment 2 were all created with the same dimension as in Experiment 1. Most of them were designed by us from scratch but only Agostini-cube was created with Stimupy. The entire experiment was composed of 60 trials: (14 stimuli + 1 filler) × 2 rotations × 2 targets. The duration was approximately 15 minutes.

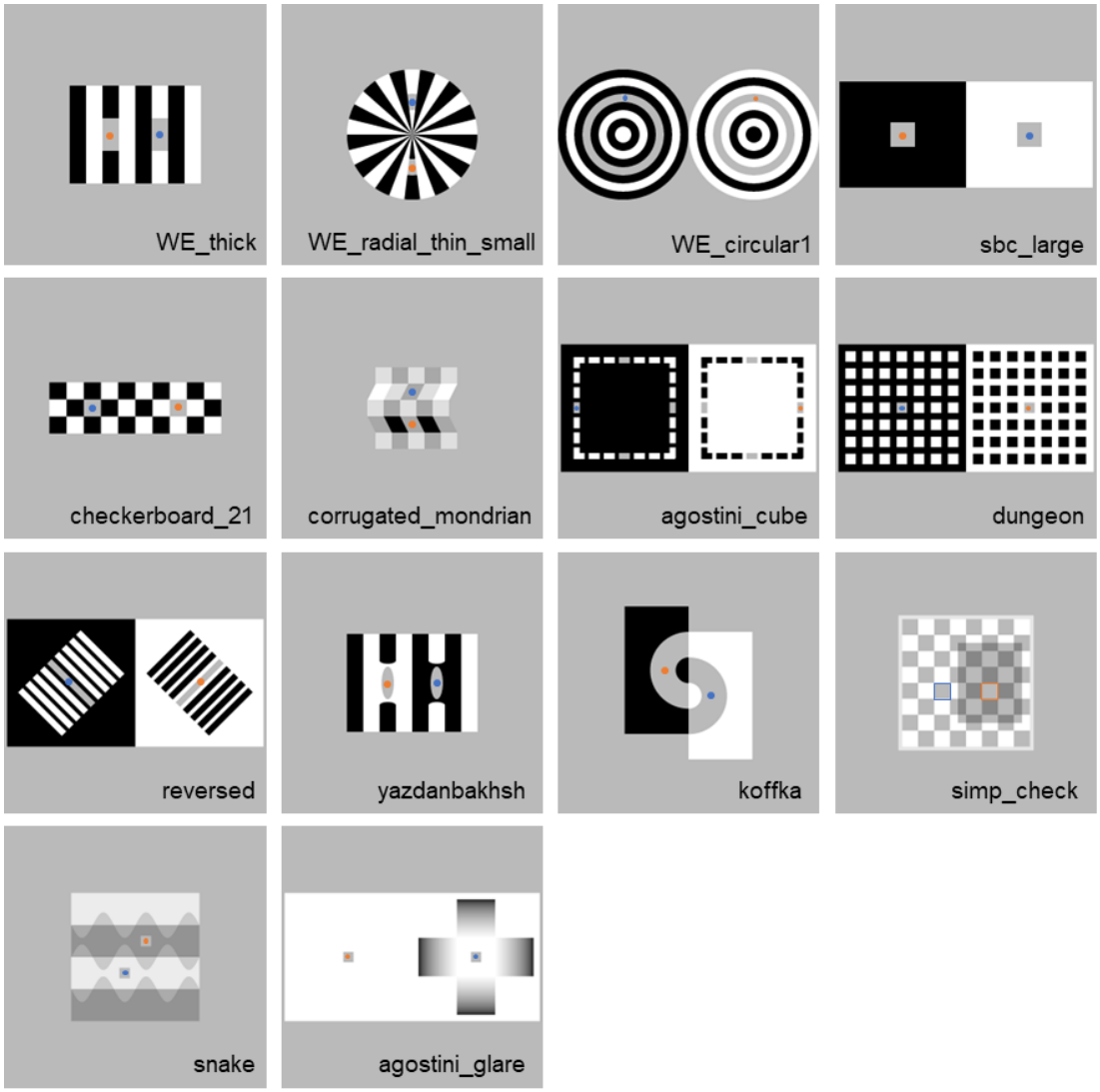

Figure 4. Fourteen illusions used in Experiment 2. Areas marked with orange are supposed to be perceived as lighter than those with blue. The first six illusions were used in Experiment 1 as well. Following are details of the newly included ones. Agostini-cube) An illusion thought to be caused by perceptual grouping between gray targets and aligned white/black rectangles (Agostini & Galmonte, 2002a). Dungeon) An illusion thought to be caused by perceptual grouping between targets and surrounding white/black squares (Bressan, 2001). Reversed) An illusion thought to be caused by perceptual grouping between targets and aligned white/black gratings (Economou et al., 2007; Gilchrist & Annan, 2002). Yazdanbakhsh) An illusion of a similar design to White's effect but without T-junction (Yazdanbakhsh et al., 2002). Koffka) Koffka's ring created by Adelson (2000). Simp-check) A simplified version of Adelson's (1995) checkershadow illusion (Blakeslee & McCourt, 2012). Snake) A contrastive illusion caused by shadow impression (Adelson, 2000). Agostini-glare) An illusion caused by luminance gradient. Immediate surroundings of the two targets are of equal luminance (white) but the remote luminance gradients darken the appearance of the central patch (Agostini & Galmonte, 2002b).

Results

Twenty-three participants were excluded with the same criteria as in Experiment 1 so the data from the remaining 118 participants were analyzed. All illusions showed significant effects in *t*-tests in the expected directions (Table 4) so no items were excluded in the following analysis. The Kaiser-Meyer-Olkin's measures of sampling adequacy were relatively high for all the data (.68) and the minimum measure in each item was .32 in Agostini_glare. $ICC_{2,k}$ in Table 4 indicates that the reliability of the measurements was generally high.

	Mean	*SD*	*t*	*p*	*ICC*
WE_thick	14.52	21.16	7.46	<.001	.64
WE_radial_thin_small	17.50	17.10	11.12	<.001	.48
WE_circular1	23.82	24.48	10.57	<.001	.77
sbc_large	11.84	21.28	6.05	<.001	.64
checkerboard_21	22.69	19.31	12.77	<.001	.69
corrugated_mondrian	19.94	20.97	10.33	<.001	.66
agostini_cube	5.82	22.44	2.82	.006	.73
dungeon	24.89	22.89	11.81	<.001	.69
reversed	13.94	23.40	6.47	<.001	.72
yazdanbakhsh	15.97	22.30	7.78	<.001	.76
koffka	7.18	22.18	3.52	.001	.65
simp_check	13.86	27.55	5.46	<.001	.70
snake	21.02	25.23	9.05	<.001	.76
agostini_glare	15.71	16.61	10.27	<.001	.62
filler	-	-	-	-	.36

Table 4. Summary of the results.

We conducted EFA on this dataset. VSS criterion, MAP criterion, and a parallel analysis suggested three, one, and two-factor models, respectively. We examined two and three-factor models (Table 5). Maximum likelihood estimation and varimax rotation were used as in Experiment 1. In the two-factor model, seven illusions showed relatively high positive loadings to Factor 1 and five illusions did on Factor 2. Koffka and Agostini-glare did not show high positive loadings to either of them. Figure 5 visualizes a two-dimensional mapping of the illusions based on the factor loadings of the two-factor model. In the three-factor model, The Factor 1 was highly loaded by seven illusions, Factor 2 was by three, and Factor 3 was by two. Corrugated-Mondrian showed weak loadings on Factors 2 and 3, and Agostini-glare did not show high loadings on any of the factors as well as in the two-factor model. In Experiment 2, the total variance the two factors accounted for was 26% (34% for the three factors).

	Factor 1	Factor 2	Comm.
WE_thick	.456	.254	.272
WE_radial_thin_small	.151	.423	.202
WE_circular1	.647	-.113	.431
sbc_large	-.175	.335	.142
checkerboard_21	.539	.055	.294
corrugated_mondrian	-.018	.468	.219
agostini_cube	.519	-.087	.277
dungeon	.708	.119	.515
reversed	.544	-.053	.298
yazdanbakhsh	.546	.205	.341
koffka	-.315	.184	.133
simp_check	-.159	.550	.328
snake	.079	.467	.224
agostini_glare	.016	.075	.006

	Factor 1	Factor 2	Factor 3	Comm.
WE_thick	.446	.014	.319	.301
WE_radial_thin_small	.125	-.001	.731	.550
WE_circular1	.647	-.145	-.129	.457
sbc_large	-.168	.182	.350	.184
checkerboard_21	.552	.043	-.029	.307
corrugated_mondrian	.003	.207	.269	.115
agostini_cube	.536	-.040	-.147	.310
dungeon	.696	-.059	.164	.515
reversed	.542	-.071	-.020	.299
yazdanbakhsh	.545	.021	.203	.338
koffka	-.270	.375	-.042	.215
simp_check	-.113	.357	.296	.228
snake	.227	.971	.016	.995
agostini_glare	.012	-.013	.089	.008

Table 5. The factor loading matrix of Experiment 2. Left is the two-factor model and right is the three factor model.

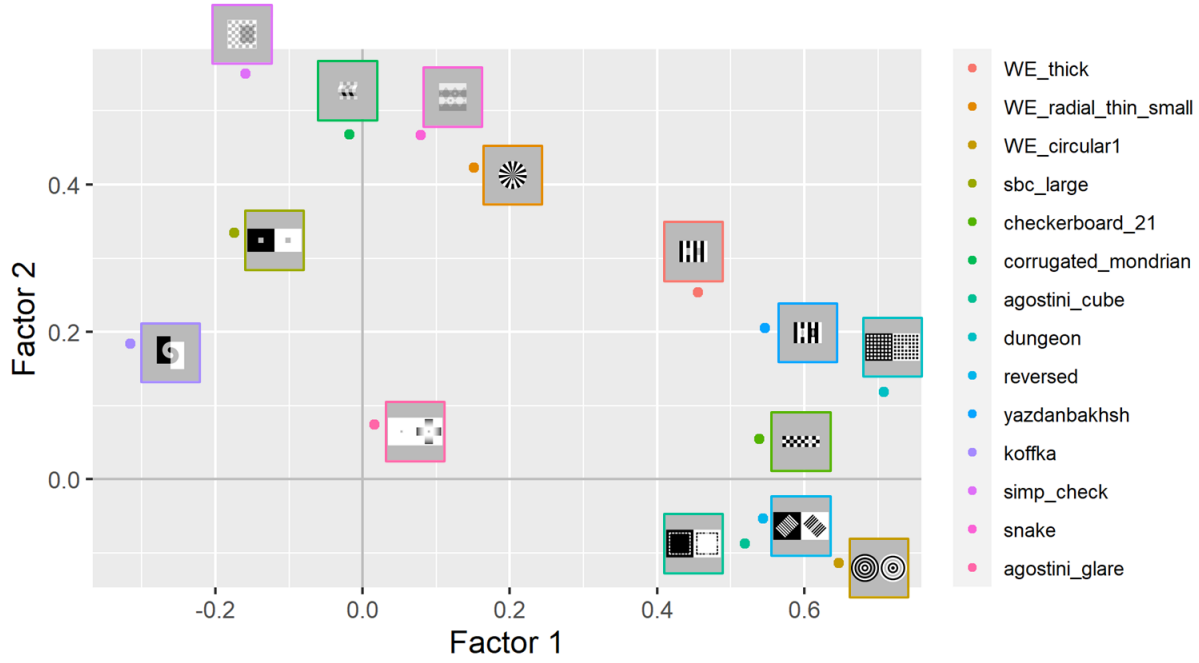

Figure 5. Two-dimensional mapping of the factor loadings in the two-factor model.

Discussions

We focus on the results of the parallel analysis in this paper and will mainly discuss the two factor model. The results roughly separated the current illusion set into contrast and assimilation classes as in the analysis of the data from Nedimović et al. (2022). Whereas reverse contrast was considered distinct from assimilation (Agostini et al., 2020), illusions that are thought to be largely driven by reverse contrast (Dungeon, Economou, and Agostini_cube) exhibited similar patterns to those associated with assimilation (e.g., Checkerboard-21). This result implies no clear difference between reverse contrast and assimilation, but the former has been considered to be caused by perceptual grouping with remote regions while the latter has

been associated with lower-level information (e.g., high spatial frequency). It is plausible that assimilation was the primary cause of the present stimuli because they were composed of repetitive patterns (Agostini et al., 2014, 2020). However, this does not deny the role of perceptual grouping in lightness perception. With an appropriate design of an illusion purely driven by grouping (Agostini et al., 2020), we would be able to observe an independent factor (Economou et al., 2015; Kaneko & Gilchrist, 2020).

In Experiment 2, we did not identify an independent factor for White's effect. In the two-factor model, WE-thick exhibited a high loading on Factor 1, which seemingly represents the factor for assimilation, and WE-radial-thin-small was more related to Factor 2, which seemed to reflect the contrast effect. However, WE-thick also showed a moderate loading (> .2) on Factor 2 unlike the other assimilative illusions, and these two illusions are plotted in relatively close proximity within the first quadrant in Figure 5. These facts partially support the notion that White's effect is not merely an aspect of contrast or assimilation phenomena. Another finding related to White's effect was that Yazdanbakhsh exhibited a similar pattern to WE-thick even though it has no T-junctions. Therefore, this illusion appears to be a more suitable example to challenge the T-junction-based explanation of White's effect than the circular variant. The evidence for the independence of White's effect is inconclusive in Experiment 2, but it offers numerous insights into the process driving this illusion, necessitating direct examination in future research.

We aimed to examine whether the illusions containing "shadowed" areas show their characteristic tendency. The specific factor for the process of shadow was not extracted in the two-factor model, but Factor 2 in the three-factor model seems to reflect it, showing high loadings from snake, Koffka, and simp-check. Considering that SBC-large exhibited its primary loading on Factor 3, Factor 2 should not be considered a factor for a simple contrast effect. It is well known that shadow-like regions enhance illusory effects (Adelson, 2000; Blakeslee & McCourt, 2012) and the present results may have shown this factor.

The observed independence of Agostini_glare was intriguing. This illusion did not show high loadings on any of the factors in both models and was not significantly correlated with any of the other illusions (see Supplement for correlation matrices). Although this illusion has not been extensively explored in model tests, its independence underscores the need to incorporate it in future model evaluations.

Model tests

In this section, we examine lightness/brightness models using the outcomes of the two experiments. We here compare three models that are based on early processes of brightness perception: ODOG, LODOG, and FLODOG (Blakeslee & McCourt, 1999; Robinson et al., 2007). They have been known as models that succeeded in explaining White's effect without explicitly processing T-junctions and depth information. They have been explored by several studies with many illusions and showed both successes and failures (Betz et al., 2015; Blakeslee & McCourt, 2004, 2012; Economou et al., 2007; Kim et al., 2018; Murray, 2020; Nedimović et al., 2022; Zeman et al., 2015). We ran these models on the illusions employed in

Experiments 1 and 2, using Multyscale, a Python package including implementations of those models (Vincent & Maertens, 2021a). Robinson et al. (2007) reported these three models' outputs for their illusion set (i.e., those used in our Experiment 1), but we again ran them on this set because 1) the reported results in Robinson et al. (2007) assumed 32 deg. × 32 deg. image size, which seems implausibly large in the environments employed in the present online experiment, and because 2) the target areas indicated in the present experiment were slightly different from those on which the reported results were based. We tested these models on the images assuming their sizes to be 16 deg. × 16 deg. The pixel values of the input images ranged from 0 to 255.

Table 6 shows the model outputs on the two illusion sets. The outputs for Experiment 1 are similar to those reported in Robinson et al. (2007) but there are slight differences. Using these outputs and the factor loadings obtained in the experiments, we calculated the scores of the models for each factor as follows. First, from a factor loading matrix Λ, Λ_{std} was defined as:

$$\Lambda_{\text{std}} = \left[\frac{a_{ij}}{\sum_{i=1}^{m} a_{ij}}\right] \text{ for } i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$$

where a_{ij} is the (i, j) entry of the m × n matrix Λ. Hence, Λ_{std} is the matrix where the sum of each of Λ's columns is normalized to be one. This normalization equalizes the weights on each factor, which can be unbalanced in a biased illusion set. Then, the factor-based model score was defined as $FBM\ Score = R^{\top}\Lambda_{\text{std}}$ using the matrix R, which binarized version of the model output with zero and one that indicates incorrect and correct predictions of illusion directions, respectively (i.e., the replacement of entries in Table 6 with 1 for red cells and 0 for blue cells). Each value in the resulting FBM score matrix can be interpreted as the percentage of each factor accounted for by the model. Theoretically, the values can exceed one due to negative values in Λ but they are unlikely in practice. The scores for the uniquenesses of the illusions can also be calculated by simply replacing Λ_{std} with the vector of the uniquenesses. Note that the scores for the common factors and those for the uniquenesses are in different scales.

Table 7 shows the models' scores calculated with the present results. The scores from Experiment 1 indicate that ODOG and LODOG are oriented toward contrast effect. FLODOG scores higher in assimilation and White's effect than ODOG and LODOG; however, its score for contrast is lower, indicating that FLODOG's superiority largely stems from its ability to predict assimilation and White's effect. A similar tendency was also observed in the scores for Experiment 2. The scores for uniqueness were roughly proportional to the simple counts of successfully predicted illusions. Based on these results, the comparative evaluation placing FLODOG over ODOG and LODOG in terms of their performances does not waver even with consideration of correlations, supporting the validity of the simple counts. However, EFA enabled the decomposition of the models' scores into those for underlying factors of illusions.

We further examined the quantitative aspect of the model outputs using the present data. Since the output from ODOG-type models is based on an arbitrary scale, it cannot be directly compared to a human response as an absolute measurement (Vincent & Maertens, 2021b).

However, a relative comparison is feasible by examining their correlations with various illusions. We conducted Kendall's rank correlation tests for the results of the two experiments respectively. Human responses were defined with the mean values of the illusion magnitudes. For Experiment 1, none of the correlations with the three model's outputs was significant (ODOG: τ_b = .089, p = .541, n = 26; LODOG: τ_b = .015, p = .930, n = 26; FLODOG: τ_b = .089, p = .541, n = 26). No significant result was found in Experiment 2, either (ODOG: τ_b = -.231, p = .280, n = 14; LODOG: τ_b = -.165, p = .451, n = 14; FLODOG: τ_b = .187, p = .388, n = 14). These tests are not conclusive since they were all post-hoc and with small sample sizes (i.e., the number of illusions), but the negative correlations in some pairs were noteworthy. Whereas those models have been known for their successes in qualitative predictions (i.e., predictions of illusion directions), the analysis of correlations indicated that their quantitative predictions cannot yet be considered satisfactory.

	ODOG	LODOG	FLODOG
WE_thick	1.00	1.00	1.00
WE_thin_wide	1.43	2.12	2.44
WE_dual	-0.24	0.74	1.24
WE_anderson	0.08	0.02	0.21
WE_zigzag	-0.57	-0.27	1.82
WE_radial_thick_small	-0.85	-0.26	0.84
WE_radial_thick	-0.60	-0.06	0.61
WE_radial_thin_small	-0.31	-0.02	2.48
WE_radial_thin	-0.21	0.14	2.46
WE_circular1	-1.15	-0.43	0.22
WE_circular05	-0.04	-0.02	1.64
grating_induction	1.58	1.74	0.99
sbc_large	5.35	2.54	2.55
sbc_small	5.99	2.91	4.31
todorovic_equal	0.94	0.33	0.07
todorovic_in_large	1.12	0.40	0.57
todorovic_in_small	1.52	0.54	1.17
todorovic_out	0.87	0.33	0.17
checkerboard_094	0.64	0.20	3.06
checkerboard_21	-0.44	-0.14	1.10
corrugated_mondrian	0.45	0.26	-0.04
benary_cross	-0.15	0.02	0.08
todorovic_benary1_2	-1.35	-0.06	-0.26
todorovic_benary3_4	1.98	0.95	0.67
bullseye_thin	3.74	3.31	0.23
bullseye_thick	3.35	1.93	2.16

	ODOG	LODOG	FLODOG
WE_thick	1.00	1.00	1.00
WE_radial_thin_small	-0.31	-0.02	2.48
WE_circular1	-1.15	-0.43	0.22
sbc_large	5.35	2.54	2.55
checkerboard_21	-0.44	-0.14	1.10
corrugated_mondrian	0.45	0.26	-0.04
agostini_cube	-4.13	-2.21	-1.74
dungeon	-1.51	-0.57	1.09
reversed	-1.67	-0.89	-1.26
yazdanbakhsh	0.23	0.34	-0.04
koffka	1.58	0.66	0.92
simp_check	0.98	0.39	0.46
snake	0.69	0.89	1.21
agostini_glare	-1.19	-0.73	-1.40

Table 6. Model outputs for all stimuli used in Experiments 1 (left) and 2 (right). Positive values (correct prediction) are indicated with orange and negative values (incorrect prediction) are with blue. Values in each table are divided by the output value for WE-thick to be standardized.

	Assimilation	White	Contrast	Uniqueness	Count
ODOG	0.59	0.42	0.78	10.89	15
LODOG	0.60	0.61	0.93	13.18	18
FLODOG	0.97	0.98	0.83	17.18	24

	Assimilation	Contrast	Uniqueness	Count
ODOG	0.12	0.85	5.34	7
LODOG	0.12	0.85	5.34	7
FLODOG	0.55	0.79	6.46	9

Table 7. Factor-based model scores and counts of correct predictions (Top: Experiment 1, Bottom: Experiment 2).

General discussions

The aim of the present study was to clarify the relationship among various lightness illusions and to enhance the methods used for evaluating computational models. The models have been evaluated with the number of illusions successfully predicted (Kobayashi & Kitaoka, 2022; Robinson et al., 2007; Zeman et al., 2015), but this naïve method assumes that each illusion is independent. We examined the correlations of the illusion magnitudes to assess the redundancy in sets of illusions and identified several underlying factors. The redundancy was not significant enough to distort the ranking of the three well-accepted models (Blakeslee & McCourt, 1999; Robinson et al., 2007). The weak redundancy suggests that there is no unique common factor underlying lightness illusions, and this is consistent with the literature that explored correlations among geometrical illusions (Cretenoud, Grzeczkowski, et al., 2020; Cretenoud et al., 2019). The identified factor structures, however, were interpretable and matched theoretical classification. They enabled us to see which class of lightness illusions are better explained by each model. The present study showed that some of the analyzed models are oriented towards contrast effects although they have been known for their successful prediction of White's effect (Blakeslee & McCourt, 1999, 2004, 2012). Moreover, an additional analysis of the rank correlations of human perceptions and model outputs revealed that the quantitative aspect of the models' performances is still behind expectations. The large dataset and multivariate analyses were highly insightful in offering detailed evaluations of the lightness models. They also contributed to understanding the relationship of various lightness illusions, although the results of EFA should not be regarded as conclusive (Mollon et al., 2017). The materials of the present study (stimuli, analysis codes, experiment programs, and data) are all available for further analyses and tests of other lightness models.

The relationship of lightness illusions revealed in the present study offered important suggestions for establishing a standard set of illusions used for model evaluation commonly (Murray, 2021). First, some lightness illusions that have been supposed to reflect the role of

perceptual grouping are largely driven by a process for assimilations. To explore the role of grouping in lightness perception, researchers need to create an illusion design purely caused by grouping. Second, White-effect-related illusions need to be included in the set since they suggested difference from contrasts and assimilations. Moreover, the contrast enhancement induced by remote luminance gradients (Agostini & Galmonte, 2002b) is highly independent of other illusions and could not be explained by the three tested models. This fact also compels a greater focus on this illusion and requires it to be used for model evaluation. Overall, the present study suggested that the illusion set should be composed of those driven by contrast, assimilation, White's effect, and contrast enhancement, and necessitated a new design for reverse contrast.

Even though the present study employed various lightness illusions, there are still many stimuli to be examined in future studies. Correlations with some new illusions that do not seem to be driven by contrast or assimilation are worth examining (Kobayashi & Morikawa, 2019, 2023). Moreover, since the present study focused on the magnitudes of illusions, we did not employ and analyze control figures, which are supposed to cause weaker or no illusory effects despite their resemblance to illusory counterparts (e.g., WE-howe in Experiment 1 was originally introduced as a weaker version of WE-anderson). Null predictions for them are also a significant point for models to satisfy. Furthermore, the use of augmented images should also be considered for testing computational models rather than relying on a single image for each illusion. Betz et al. (2015) tested ODOG model using multiple noise-masked versions of White's illusion. Although the illusion set used in Experiment 1 contained some image pairs that can be considered size-augmented versions of the others (e.g., WE-circular 1 and WE-circular 05), other attributes of images, such as luminance, width-height ratio, or rotation can also be modified and may have a significant impact on models' responses. Models' robustness to image augmentation is an important evaluation metric. A wider variety of images will benefit the further deepening of understanding of lightness illusions and models.

Since the present study employed online experiments, some portion of the variance in the datasets may have been introduced by the variety of observation environments. Especially, differences of visual extents should differentiate frequencies of patterns in some illusions, and this difference may have influenced the magnitudes and directions of those illusions. Based on the fact that a higher frequency of a repetitive pattern (per visual degree) enhances assimilation (Blakeslee & McCourt, 2004), the loadings of the assimilation factor might have been moderately inflated due to the potential variety in pattern frequencies. For further examination of the relationships of the lightness illusions, controlling visual extents in online environments (Li et al., 2020) can be a key point.

It is desired to establish a computational model capable of predicting human lightness perception. With the development of several models, evaluation of them becomes as important as inventing a new model, but insufficient attention has been paid to how we should evaluate them. In this study, we clarified the similarities and differences of lightness illusions using large datasets of human responses. They served as useful information to examine models in greater detail and offer insightful suggestions for further sophistication of evaluations. The study highlighted the need to consider underlying processes rather than focusing solely on

individual illusions, and the present findings will lead to a rigorous theory of human lightness perception.

Acknowledgment

This study was funded by the Japan Society for the Promotion of Science (grant number: 22K13878). ChatGPT was used for proofreading in the preparation of the draft.

The authors would like to thank Dr. Michael Herzog for his comments on the draft of this paper, which greatly improved its quality.

References

Adams, W. J. (2007). A common light-prior for visual search, shape, and reflectance judgments. *Journal of Vision*, *7*(11), 11. https://doi.org/10.1167/7.11.11

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, *262*(5142), 2042–2044. https://doi.org/10.1126/science.8266102

Adelson, E. H. (2000). Lightness Perception and Lightness Illusions. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (pp. 339–351). MIT Press.

Agostini, T., & Galmonte, A. (2002a). Perceptual Organization Overcomes the Effects of Local Surround in Determining Simultaneous Lightness Contrast. *Psychological Science*, *13*(1), 89–93. https://doi.org/10.1111/1467-9280.00417

Agostini, T., & Galmonte, A. (2002b). A new effect of luminance gradient on achromatic simultaneous contrast. *Psychonomic Bulletin & Review*, *9*(2), 264–269. https://doi.org/10.3758/BF03196281

Agostini, T., Murgia, M., & Galmonte, A. (2014). Reversing the Reversed Contrast. *Perception*, *43*(2–3), 207–213. https://doi.org/10.1068/p7553

Agostini, T., Murgia, M., Sors, F., Prpic, V., & Galmonte, A. (2020). Contrasting a Misinterpretation of the Reverse Contrast. *Vision (Basel, Switzerland)*, *4*(4). https://doi.org/10.3390/vision4040047

Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. *Perception*, *26*(4), 419–453. https://doi.org/10.1068/p260419

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

Bindman, D., & Chubb, C. (2004). Brightness assimilation in bullseye displays. *Vision Research*, *44*(3), 309–319. https://doi.org/10.1016/S0042-6989(03)00430-9

Blakeslee, B., & McCourt, M. E. (1999). A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. *Vision Research*, *39*(26), 4361–4377. https://doi.org/10.1016/S0042-6989(99)00119-4

Blakeslee, B., & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Research*, *44*(21), 2483–2503.

Blakeslee, B., & McCourt, M. E. (2012). When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision Research*, *60*, 40–50. https://doi.org/10.1016/j.visres.2012.03.006

Blakeslee, B., Padmanabhan, G., & McCourt, M. E. (2016). Dissecting the influence of the collinear and flanking bars in White's effect. *Vision Research*, *127*, 11–17. https://doi.org/10.1016/j.visres.2016.07.001

Bressan, P. (2001). Explaining Lightness Illusions. *Perception*, *30*(9), 1031–1046. https://doi.org/10.1068/p3109

Bressan, P. (2006). The place of white in a world of grays: A double-anchoring theory of lightness perception. *Psychological Review*, *113*(3), 526–553. https://doi.org/10.1037/0033-295X.113.3.526

Clifford, C. W. G., & Spehar, B. (2003). Using colour to disambiguate contrast and assimilation in White's Effect. *Journal of Vision*, *3*(9), 294–294. https://doi.org/10.1167/3.9.294

Cretenoud, A. F., Francis, G., & Herzog, M. H. (2020). When illusions merge. *Journal of Vision*, *20*(8), 12–12. https://doi.org/10.1167/jov.20.8.12

Cretenoud, A. F., Grzeczkowski, L., Bertamini, M., & Herzog, M. H. (2020). Individual differences in the Müller-Lyer and Ponzo illusions are stable across different contexts. *Journal of Vision*, *20*(6), 4–4. https://doi.org/10.1167/jov.20.6.4

Cretenoud, A. F., Karimpur, H., Grzeczkowski, L., Francis, G., Hamburger, K., & Herzog, M. H. (2019). Factors underlying visual illusions are illusion-specific but not feature-specific. *Journal of Vision*, *19*(14), 12–12. https://doi.org/10.1167/19.14.12

DeValois, R. L., & DeValois, K. K. (1990). *Spatial Vision*. Oxford University Press. https://play.google.com/store/books/details?id=_fO6wzoVK2wC

Economou, E., Zdravkovic, S., & Gilchrist, A. L. (2007). Anchoring versus spatial filtering accounts of simultaneous lightness contrast. *Journal of Vision*, *7*(12), 2. https://doi.org/10.1167/7.12.2

Economou, E., Zdravković, S., & Gilchrist, A. L. (2015). Grouping Factors and the Reverse Contrast Illusion. *Perception*, *44*(12), 1383–1399. https://doi.org/10.1177/0301006615607118

Gilchrist, A. L. (2006). *Seeing black and white*. Oxford University Press.

Gilchrist, A. L., & Annan, V. (2002). Articulation Effects in Lightness: Historical Background and Theoretical Implications. *Perception*, *31*(2), 141–150. https://doi.org/10.1068/p04sp

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

Gordillo, D., Cretenoud, A., Garobbio, S., & Herzog, M. H. (2023). A solution to the ill-posed problem of common factors in vision. *Journal of Vision*, *23*(9), 5081–5081. https://doi.org/10.1167/jov.23.9.5081

Grzeczkowski, L., Clarke, A. M., Francis, G., Mast, F. W., & Herzog, M. H. (2017). About individual differences in vision. *Vision Research*, *141*, 282–292. https://doi.org/10.1016/j.visres.2016.10.006

Grzeczkowski, L., Cretenoud, A., Herzog, M. H., & Mast, F. W. (2017). Perceptual learning is specific beyond vision and decision making. *Journal of Vision*, *17*(6), 6. https://doi.org/10.1167/17.6.6

Howe, P. D. L. (2005). White's Effect: Removing the Junctions but Preserving the Strength of the Illusion. *Perception*, *34*(5), 557–564. https://doi.org/10.1068/p5414

Kaneko, S., & Gilchrist, A. (2020). Lightness in a Flash: Effect of Exposure Time on Lightness Perception. *I-Perception*, *11*(6), 2041669520983830. https://doi.org/10.1177/2041669520983830

Kaneko, S., Murakami, I., Kuriki, I., & Peterzell, D. H. (2018). Individual Variability in Simultaneous Contrast for Color and Brightness: Small Sample Factor Analyses Reveal Separate Induction Processes for Short and Long Flashes. *I-Perception*, *9*(5), 204166951880050. https://doi.org/10.1177/2041669518800507

Kim, M., Gold, J. M., & Murray, R. F. (2018). What image features guide lightness perception? *Journal of Vision*, *18*(13), 1. https://doi.org/10.1167/18.13.1

Kobayashi, Y., & Kitaoka, A. (2022). Simple Assumptions to Improve Markov Illuminance and Reflectance. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.915672

Kobayashi, Y., & Morikawa, K. (2019). An Upward-Facing Surface Appears Darker: The Role Played by the Light-From-Above Assumption in Lightness Perception. *Perception*, *48*(6), 500–514. https://doi.org/10.1177/0301006619847590

Kobayashi, Y., & Morikawa, K. (2023). Vertical anisotropy in lightness perception not caused by lighting assumption. *Vision Research*, *206*, 108193. https://doi.org/10.1016/j.visres.2023.108193

Kobayashi, Y., Zavagno, D., & Morikawa, K. (2021a). Distinct processes of lighting priors for lightness and 3-D shape perception. *Journal of Vision*, *21*(6), 1–1. https://doi.org/10.1167/jov.21.6.1

Kobayashi, Y., Zavagno, D., & Morikawa, K. (2021b). Asymmetric Brightness Effects With Dark Versus Light Glare-Like Stimuli. *I-Perception*, *12*(1), 204166952199314. https://doi.org/10.1177/2041669521993144

Land, E. H., & McCann, J. J. (1971). Lightness and Retinex Theory. *Journal of the Optical Society of America*, *61*(1), 1–11. https://doi.org/10.1364/JOSA.61.000001

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for Participants' Viewing Distance in Large-Scale, Psychophysical Online Experiments Using a Virtual Chinrest. *Scientific Reports*, *10*(1), 904. https://doi.org/10.1038/s41598-019-57204-1

McCourt, M. E. (1982). A spatial frequency dependent grating-induction effect. *Vision Research*, *22*(1), 119–134. https://doi.org/10.1016/0042-6989(82)90173-0

Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Individual Differences as a Window into the Structure and Function of the Visual System*, *141*, 4–15. https://doi.org/10.1016/j.visres.2017.11.001

Murray, R. F. (2020). A model of lightness perception guided by probabilistic assumptions about lighting and reflectance. *Journal of Vision*, *20*(7), 28. https://doi.org/10.1167/jov.20.7.28

Murray, R. F. (2021). Lightness Perception in Complex Scenes. *Annual Review of Vision Science*, *7*(1), 417–436. https://doi.org/10.1146/annurev-vision-093019-115159

Nedimović, P., Zdravković, S., & Domijan, D. (2022). Empirical evaluation of computational models of lightness perception. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-22395-7

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peterzell, D. H. (2016). Discovering Sensory Processes Using Individual Differences: A Review and Factor Analytic Manifesto. *Electronic Imaging*, *2016*(16), 1–11. https://doi.org/10.2352/ISSN.2470-1173.2016.16.HVEI-112

Peterzell, D. H., Serrano-Pedraza, I., Widdall, M., & Read, J. C. A. (2017). Thresholds for sine-wave corrugations defined by binocular disparity in random dot stereograms: Factor

analysis of individual differences reveals two stereoscopic mechanisms tuned for spatial frequency. *Vision Research*, *141*, 127–135. https://doi.org/10.1016/j.visres.2017.11.002

Peterzell, D. H., Werner, J. S., & Kaplan, P. S. (1995). Individual differences in contrast sensitivity functions: longitudinal study of 4-, 6- and 8-month-old human infants. *Vision Research*, *35*(7), 961–979. https://doi.org/10.1016/0042-6989(94)00117-5

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. https://doi.org/10.1016/j.visres.2007.02.017

Rudd, M. E. (2010). How attention and contrast gain control interact to regulate lightness contrast and assimilation: A computational neural model. *Journal of Vision*, *10*(14), 40–40. https://doi.org/10.1167/10.14.40

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

Shaqiri, A., Pilz, K. S., Cretenoud, A. F., Neumann, K., Clarke, A., Kunchulia, M., & Herzog, M. H. (2019). No evidence for a common factor underlying visual abilities in healthy older people. *Developmental Psychology*, *55*(8), 1775–1787. https://doi.org/10.1037/dev0000740

Takahashi, K., Ujiie, Y., & Dai, Z. (2023, August). *Inter-Individual Variability of Visual Illusions: Validity and Potential of Online Experiments as a Tool for Visual Illusion Study*. European Conference on Visual Perception, Paphos.

Taya, R., Ehrenstein, W. H., & Cavonius, C. R. (1995). Varying the Strength of the Munker—White Effect by Stereoscopic Viewing. *Perception*, *24*(6), 685–694. https://doi.org/10.1068/p240685

Todorović, D. (1997). Lightness and Junctions. *Perception*, *26*(4), 379–394. https://doi.org/10.1068/p260379

Vincent, J., & Maertens, M. (2021a). A history and modular future of multiscale spatial filtering models. *Journal of Vision*, *21*(9), 2824–2824. https://doi.org/10.1167/jov.21.9.2824

Vincent, J., & Maertens, M. (2021b). *The missing linking functions in computational models of brightness perception* (Vol. 50). Sage Publications Ltd.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, *38*(3), 310−324.

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

White, M. (1981). The Effect of the Nature of the Surround on the Perceived Lightness of Grey Bars within Square-Wave Test Gratings. *Perception*, *10*(2), 215–230. https://doi.org/10.1068/p100215

Yazdanbakhsh, A., Arabzadeh, E., Babadi, B., & Fazl, A. (2002). Munker–White-Like Illusions without T-Junctions. *Perception*, *31*(6), 711–715. https://doi.org/10.1068/p3348

Yund, E. W., & Armington, J. C. (1975). Color and brightness contrast effects as a function of spatial variables. *Vision Research*, *15*(8), 917–929. https://doi.org/10.1016/0042-6989(75)90231-X

Zavagno, D., Daneyko, O., & Liu, Z. (2018). The Influence of Physical Illumination on Lightness Perception in Simultaneous Contrast Displays. *I-Perception*, *9*(4), 204166951878721. https://doi.org/10.1177/2041669518787212

Zeman, A., Brooks, K. R., & Ghebreab, S. (2015). An exponential filter model predicts lightness illusions. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00368