

# Conjoint measurement of perceived transparency and perceived contrast in variegated checkerboards

Guillermo Aguilar

Computational Psychology, Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany



Marianne Maertens

Computational Psychology, Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany



One fundamental question in vision research is how the retinal input is segmented into perceptually relevant variables. A striking example of this segmentation process is transparency perception, in which luminance information in one location contributes to two perceptual variables: the properties of the transparent medium itself and of what is being seen in the background. Previous work by Robilotto et al. (2002, 2004) suggested that perceived transparency is closely related to perceived contrast, but how these two relate to retinal luminance has not been established. Here we studied the relationship between perceived transparency, perceived contrast, and image luminance using maximum likelihood conjoint measurement (MLCM). Stimuli were rendered images of variegated checkerboards that were composed of multiple reflectances and partially covered by a transparent overlay. We systematically varied the transmittance and reflectance of the transparent medium and measured perceptual scales of perceived transparency. We also measured scales of perceived contrast using cut-outs of the transparency stimuli that did not contain any geometrical cues to transparency. Perceptual scales for perceived transparency and contrast followed a remarkably similar pattern across observers. We tested the empirically observed scales against predictions from various contrast metrics and found that perceived transparency and perceived contrast were equally well predicted by a metric based on the logarithm of Michelson or Whittle contrast. We conclude that judgments of perceived transparency and perceived contrast are likely to be supported by a common mechanism, which can be computationally captured as a logarithmic contrast.

## Introduction

The human visual system senses light and processes the sensory input such that the owner of the visual

system, the human observer, perceives meaningful information about the environment. Considering a single fixation, the sensory input to the retina is often likened to a static image of the world, which is highly ambiguous with respect to the elements in the viewed scene. For example, when part of the scene is covered by a transparent medium (Figure 1A), the stimulation at one retinal location results from two sources, the light reflected from the transparent medium and the light reflected from the surface behind the transparency. Phenomenologically, human observers clearly perceive both sources separately. The perceptual process that assigns meaningful perceptual categories to the sensory input is called perceptual segmentation. To date, there is no model that would take an image as input and predict the corresponding percepts as output.

Prior work suggests that it is the (perceived) contrast in the regions of transparency and in plain view that provides the crucial link between image intensity and perceived transparency<sup>1</sup> (Singh & Anderson, 2002, 2006; Robilotto et al., 2002; Anderson et al., 2006; Anderson, 2008; Wiebel et al., 2017). The dependency of perceived transparency on contrast is illustrated in Figure 1A. Most observers perceive the light transparent medium as less transparent than the dark one, although both of them have the same physical transmittance.

The dependency of perceived transparency on the reflectance of the transparent medium is at odds with physical models of transparency. For example, in the episotister model by Metelli (Figure 1B) (Metelli, 1970, 1974), transparency exclusively depends on the transmittance of the medium. We introduce the episotister model to illustrate this issue and to introduce the terms that we will use in this article. The episotister is a disc with a sector that rotates in front of a background. At appropriate rotation speed, the disc appears transparent and the degree of perceived

Citation: Aguilar, G., & Maertens, M. (2022). Conjoint measurement of perceived transparency and perceived contrast in variegated checkerboards. *Journal of Vision*, 22(2):2, 1–13, <https://doi.org/10.1167/jov.22.2.2>.



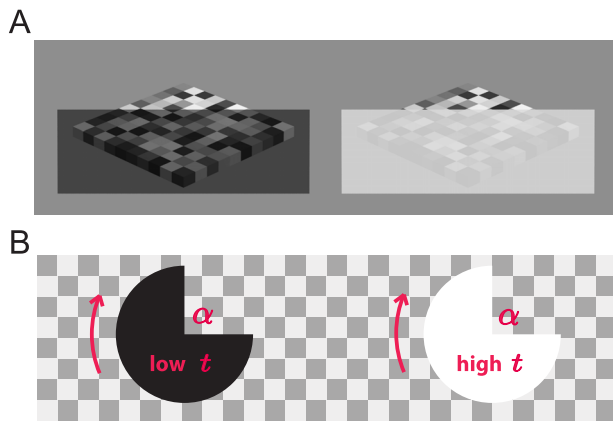


Figure 1. (A) Two transparencies rendered with different transmittance, otherwise identical in transmittance ( $\alpha$ ) and luminance range in transparency. (B) Two episcotisters with equal opening  $\alpha$  that, when rotated at high speed, will have equal physical transmittance. To an observer, the black one (low  $t$ ) appears more transparent than the white one (high  $t$ ).

transparency depends on the sector size  $\alpha$ . The model predicts the luminances of the surfaces seen through the transparency  $T$  as follows:

$$T = \alpha \cdot P + (1 - \alpha) \cdot t \quad (1)$$

whereby  $P$  is the luminance of the surface seen in plain view,  $t$  is the luminance reflected from the rotating disk of a given reflectance, and  $\alpha$  is the sector. Formulating two such equations, one for the surfaces with the highest (max) and one for the surfaces with the lowest reflectance (min), one can solve them for  $\alpha$ . In the model, transmittance depends on the ratio of the luminance range in transparency to the luminance range in plain view:

$$\alpha = \frac{T_{max} - T_{min}}{P_{max} - P_{min}} \quad (2)$$

However, as illustrated in Figure 1, perceived transmittance also depends on the reflectance of the transparent medium that is not captured by the ratio of luminance ranges.

Singh and Anderson (2002, 2006; Anderson et al., 2006; Anderson, 2008) showed that in simple stimuli, which consisted of only two surfaces, perceived transparency was well predicted by a ratio of contrasts:

$$\alpha_c = \frac{C_{TRANSP}}{C_{PLAIN}} \quad (3)$$

where  $C_{TRANSP}$  is the contrast of the region in transparency and  $C_{PLAIN}$  is the contrast in plain

view (Singh & Anderson, 2006, 2002). Contrast was calculated as a Michelson contrast, defined as

$$C = \frac{l_{max} - l_{min}}{l_{max} + l_{min}} \quad (4)$$

with  $l_{min}$  and  $l_{max}$  being the minimum and maximum luminances in each region, respectively.

Similarly to Singh and Anderson (2002, 2006), Kasrai and Kingdom (2001) also reported that observers' adjustments of perceived transparency could be predicted by the ratio of Michelson contrasts. Using a stimulus array of six instead of only four different luminance values in plain view and the region of transparency, they could derive distinct predictions for the episcotister and the ratio of contrast model. Practically, however, given the luminance values for their set of stimuli, the differences between model predictions were rather small and thus limited the power to perform model selection (Kasrai & Kingdom, 2001).

Natural scenes are composed of multiple surfaces, which raises the question of whether the perception of transparency can still be captured by a contrast metric that takes only the minimum and maximum, or a limited number of luminances of the scene into account. To address that question, Robilotto and colleagues (Robilotto et al., 2002; Robilotto & Zaidi, 2004) measured perceived transparency for achromatic stimuli consisting of randomly oriented, randomly sized, and partially overlapping ellipses of varying luminances ("dead leaves"). Observers were shown two transparent overlays that moved in front of the ellipses. They adjusted the test so that it matched the perceived transmittance of the target. Observers also matched the apparent contrast of the overlay regions when all cues to transparency were removed. The settings for perceived transparency and perceived contrast were very similar, so the authors concluded that perceived contrast is the sensory determinant of perceived transparency also for scenes with multiple surfaces. In the next step, the authors used various contrast metrics (Michelson contrast, root mean square [RMS] contrast, and Whittle contrast) to predict perceived transparency and perceived contrast. None of these metrics predicted appearance for the dead leaves stimuli.

Robilotto et al. (2002) used an asymmetric matching task to measure perceived contrast and perceived transparency as a function of physical transparency parameters. While adjustment procedures are very efficient and thus widely used in the assessment of appearance, they probe the perceptual dimension of interest only indirectly (Kingdom & Prins, 2010; Wiebel et al., 2017; Aguilar & Maertens, 2020). Hence, matching procedures are insufficient to differentiate between potentially different transducer functions relating sensory to perceived variables (we discuss the

linking assumptions in matching procedures in detail below).

New scaling methods such as maximum likelihood conjoint measurement (MLCM; Knoblauch & Maloney, 2012) require more trials but offer a number of advantages compared to the matching procedure when measuring appearance. MLCM provides an estimate of the entire perceptual scale. It does so by modeling observers' appearance judgments within a statistical framework. MLCM finds the estimates of perceptual scales by maximizing the likelihood for a set of scale values given the responses of observers across all trials. Analogous to classical signal detection theory, MLCM has a stochastic component to explicitly model that observer responses can be noisy. Furthermore, MLCM allows straightforward model comparison using statistical tests (a detailed review of MLCM can be found in Maloney & Knoblauch, 2020). Here we used MLCM to measure perceptual scales of perceived transparency and perceived contrast and evaluated them against the quantitative predictions from different contrast metrics.

To anticipate, we found that perceptual scales for perceived transparency and for perceived contrast depend on both the transmittance and reflectance of the transparent medium. Consistent with previous reports (e.g., Singh & Anderson, 2002; Robilotto et al., 2002), both perceptual attributes decrease with increasing luminance in a similar way. Different from prior reports (Robilotto et al., 2002), we found that perceived transparency and perceived contrast scales

were reasonably well predicted by the space-averaged logarithm of Michelson or Whittle contrasts (Moulden et al., 1990). Our results suggest that these metrics computationally characterize the internal dimension that is relevant for judgments of perceived transparency and contrast.

## Method

### Observers

Seven observers participated in the study; two were the authors (O1, O5), one was an experienced observer (O2), and the other four were volunteers naive to the purpose of the experiment. All observers had normal or corrected to normal visual ability. Naive observers were reimbursed for participation. Informed written consent was given by all observers prior to the experiment.

### Stimuli

*Variegated checkerboards.* We used *povray* (Persistence of Vision Raytracer Pty. Ltd., Williamstown, Victoria, Australia, 2004) to render images of variegated checkerboards that were composed of  $8 \times 8$  checks (Figure 2A). The positions of the checkerboard, the light source, and the camera were kept constant across all images. Each check had one out of 13 possible reflectance values that were randomly assigned except

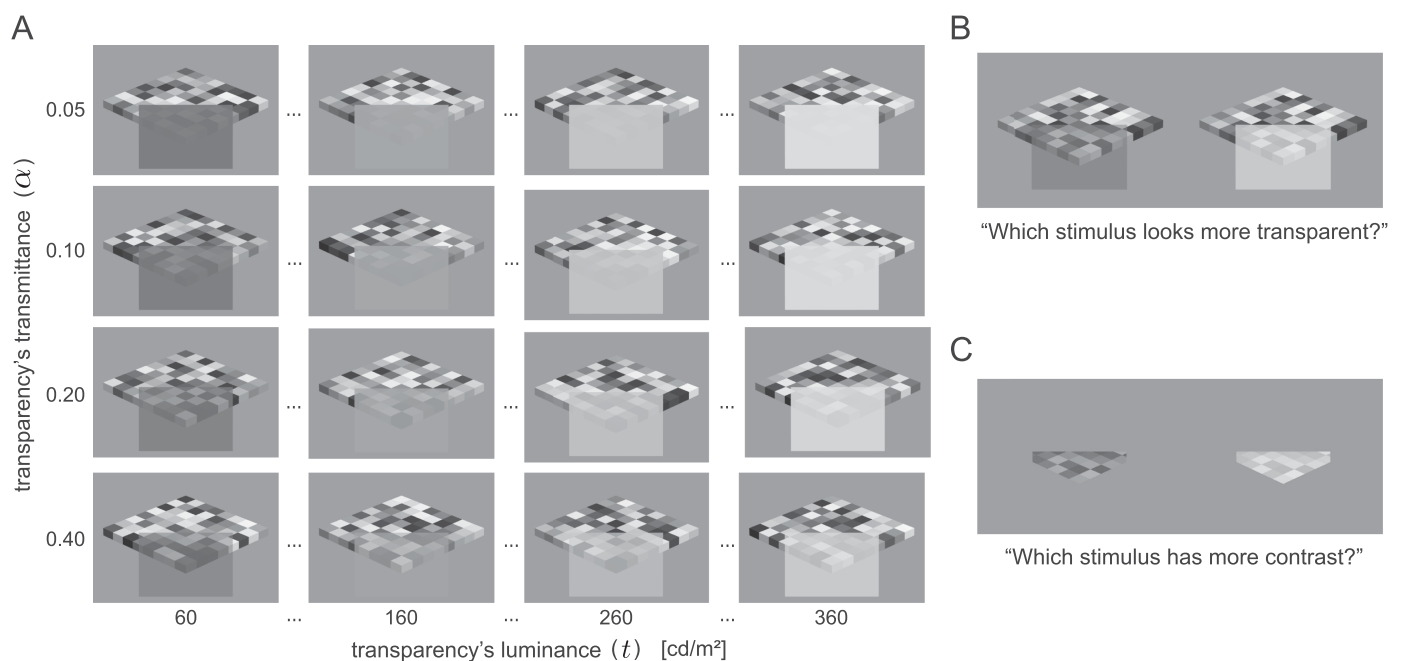


Figure 2. (A) Example set of stimuli and experimental task. The transparency's transmittance ( $\alpha$ ) and luminance when opaque ( $t$ ) were varied in  $4 \times 9$  possible combinations. (B) In Experiment 1, the observer judged which of two transparencies look more transparent. (C) In Experiment 2, the observer judged which of the two cut-outs had higher contrast.

that no two adjacent checks had the same reflectance. The luminances in “plain view”—the region without transparency—ranged from 12 to 412 cd/m<sup>2</sup>.

A transparent layer was placed between the checkerboard and the camera (Figure 2A). We varied two parameters of the transparent layer—its transmittance  $\alpha$  and its luminance  $t$  that would result for a transmittance value of zero (i.e., an opaque surface). The latter was manipulated by varying the transparent layer’s reflectance in the rendering software.

We rendered all possible combinations of four values of  $\alpha$  (0.05, 0.1, 0.2, 0.4) and nine values of  $t$  (ranging from 60 to 360 cd/m<sup>2</sup>). The background luminance was 140 cd/m<sup>2</sup>. A complete listing of luminance values and ranges separated for each condition can be found in Supplementary Table S1. The transparency was rendered using alpha blending, which implements the episcotister model (see Equation 1). A different checkerboard was rendered for each trial and for each observer.

*Cut-out stimuli.* To create stimuli without cues to depth or transparency but identical to the original stimuli with respect to luminance, we cut out the regions of transparency from the variegated checkerboards and presented them in isolation (Figure 2C). Luminance, contrast, and background luminance values were thus identical to those in the transparency region of variegated checkerboards.

## Apparatus

Stimuli were presented on a 21-in. Siemens SMM2106LS monitor (400 × 300 mm, 1,024 × 768 px, 130 Hz). Monitor calibration was conducted using a Minolta LS-100 photometer (Konica Minolta, Tokyo, Japan). The monitor’s gamma function was measured along its input range at 8-bit precision. With these measurements, we linearized the monitor luminance response over its entire output range (up to 480 cd/m<sup>2</sup>). The accuracy of linearization was estimated by repeating the luminance measurement three times and computing the standard deviation of the three samples measured at the same nominal input divided by their mean. The maximum was 0.9%.

Observers were seated 130 cm away from the screen in a dark experimental cabin. Presentation was controlled by a DataPixx toolbox (Vpixx Technologies, Inc., Saint-Bruno, QC, Canada) and custom presentation software (<http://github.com/computational-psychology/hrl>). Observers’ responses were registered with a ResponsePixx button-box (VPixxTechnologies, Inc.).

## Design and procedure

We measured perceptual scales for perceived transparency (Experiment 1) and for perceived contrast

(Experiment 2) using MLCM and the method of paired comparisons (Knoblauch & Maloney, 2012).

In each trial, two stimuli were presented simultaneously (Figures 2B and C). In Experiment 1, observers judged which of two transparent overlays looked more transparent (Figure 2B). In Experiment 2, observers judged which of the two cut-out stimuli had higher contrast (Figure 2C). To indicate their response, observers pressed the left or right button on the response box. There was no time limit.

We used 36 stimulus combinations (four levels of transparency’s transmittance  $\alpha$  and nine levels of its luminance  $t$ ), which resulted in 630 (36 · (36 – 1)/2) unique paired comparisons excluding pairs with identical stimuli. We repeated each comparison 10 times, resulting in a total of 6,300 trials per observer and experiment. The stimulus placement (left or right) and trial order were randomized in Experiment 1. Trials were divided in 10 blocks of 630 trials, which lasted approximately 30 min each. Experiment 2 used the same trial sequence as in Experiment 1.

Observers completed both experiments in multiple sessions of 1 to 2 h, including breaks. They were free to choose how many blocks they wanted to complete per session. The order of experiments was fixed: First they completed Experiment 1 and then Experiment 2.

## Scale estimation

MLCM allows one to estimate scales for a perceptual dimension as a function of two or more stimulus dimensions (Knoblauch & Maloney, 2012). In our design, the stimulus dimensions were the transparency’s transmittance ( $\alpha$ ) and its luminance ( $t$ , Equation 1). The perceptual dimensions were perceived transparency in Experiment 1 and perceived contrast in Experiment 2.

MLCM provides three decision models that can be fitted to observers’ data; we considered all of them (as suggested in Knoblauch & Maloney, 2012). The first is the “independent” model, in which observer responses are modeled to depend only on one of the stimulus dimensions,  $\alpha$  (or  $t$ ). It is the most constrained model, and for our design, it had three (or eight) free parameters. The second model is the “additive” model, which models observer responses to depend on a sum of the effects from each stimulus dimension. This model is the stochastic analogy to classical additive conjoint measurement (Luce & Tukey, 1964). In our design, this model had 11 free parameters. Finally, the “saturated” model is the most general model and fits every combination of  $\alpha$  and  $t$  as a separate effect. The model had the maximum number of possible free parameters, that is, 35 (and thus the name “saturated”). The three models were nested: The saturated was the most general one, followed by the additive model, and finally the independent model. Here we fitted all three models to our observers’ data and tested which one

explains the data best by means of a nested likelihood ratio test (Knoblauch & Maloney, 2012).

We used the MLCM implementation available for the R programming language (R Core Team, 2017; Knoblauch & Maloney, 2014). This implementation uses a generalized linear model (GLM) to estimate the scale parameters, and it readily provides routines for the calculation of confidence intervals by means of bootstrap techniques and for the evaluation of goodness of fit by means of Monte Carlo simulations. Knoblauch and Maloney (2012, 2014) provide detailed information about these statistical procedures; a summary can be found in the Supplementary Material.

## Results

### Perceptual scales

Figure 3A shows perceptual scales of perceived transparency as a function of transmittance  $\alpha$  (x-axis) and luminance  $t$  (color-coded) for one observer. Data were fitted with MLCM's "additive" model to capture the contributions of the two variables (see Method). Although the more general, "saturated" model provided a better fit in almost all cases (except for O1 in Experiment 2), the differences in perceptual scales were minimal and did not affect the pattern of results (see Supplementary Material for a detailed discussion of this point).

We observe two main effects. First, perceived transparency increases with increasing physical transmittance (positive slope), and second, perceived transparency decreases with increasing luminance (orderly color scale between functions). The second effect quantifies the phenomenological observation that

a dark transparent medium looks more transparent than a light one for identical physical transmittance (Tudor-Hart, 1928). This is consistent with previous work (e.g., Singh & Anderson, 2002; Robilotto et al., 2002), but the scales show that this dependence on the luminance of the transparent medium is maintained across the range of transmittances that we tested here.

To better illustrate the effect of the transparency's luminance  $t$  on perceived transparency, we replot the perceptual scales from Figure 3 as a function of  $t$  and plot different curves for different  $\alpha$ . Figure 4 shows that the pattern of results is consistent across individual observers. The scales show that two different stimuli with different combinations of transmittance and luminance will be perceived as equally transparent. The dashed red line in Figure 4 (left top panel) illustrates such a "trade-off" between transmittance and luminance for one observer (O1). At a perceptual scale value of 10, a stimulus with  $\alpha = 0.1$  and luminance  $t = 110 \text{ cd/m}^2$  would look equally transparent to a stimulus with  $\alpha = 0.2$  and  $t = 234 \text{ cd/m}^2$ .

Figures 3 and 4 also show perceptual scales for perceived contrast. Comparison of columns A and B shows that the scales for perceived contrast are similar to those of perceived transparency. Pearson's correlation coefficients between scales for perceived transparency (Experiment 1) and perceived contrast (Experiment 2) range from 0.975 to 0.997 (average 0.99) across observers (Supplementary Figure S2 shows their comparison).

We evaluated goodness of fit for the perceptual scales using the routines provided by the MLCM software package and as suggested by Knoblauch and Maloney (2012). In both experiments, the initial goodness of fit was insufficient for all but two cases (14%, O7 in Experiment 1 and O1 in Experiment 2). We evaluated the influence of "outliers" in the data (i.e., responses due to lapses of attention or fingerslips). We define an outlier as a data point for which the deviance residuals were higher than a threshold set by visual inspection of their distribution, as suggested by Knoblauch and Maloney (2012) (see Supplementary Information for details). The number of trials classified as outliers ranged from 19 to 32 out of 6,300 trials in individual observers (0.3% to 0.5%). Removing these trials did not affect the pattern of results. After outlier removal, all perceptual scales passed the goodness-of-fit test (Supplementary Table S2).

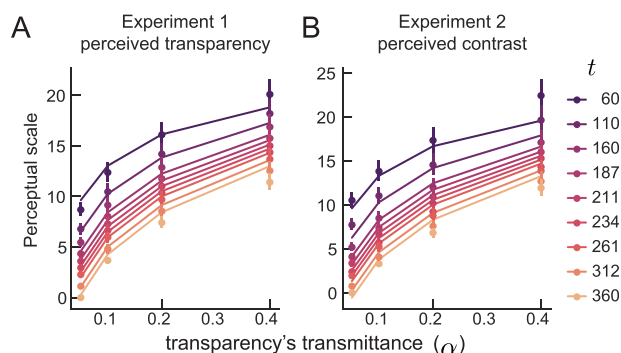


Figure 3. Perceptual scales obtained by MLCM for perceived transparency (A) and perceived contrast (B) for one observer (O1). Markers indicate scale values; their error bars indicate their 95% confidence intervals. Continuous lines depict the prediction from the space-averaged logarithm of the Michelson contrast (SAMLG).

### Contrast metrics

Following the approach by Robilotto et al. (2002), we calculated various contrast metrics for our stimuli to test which metric could quantitatively account for the empirical scales.

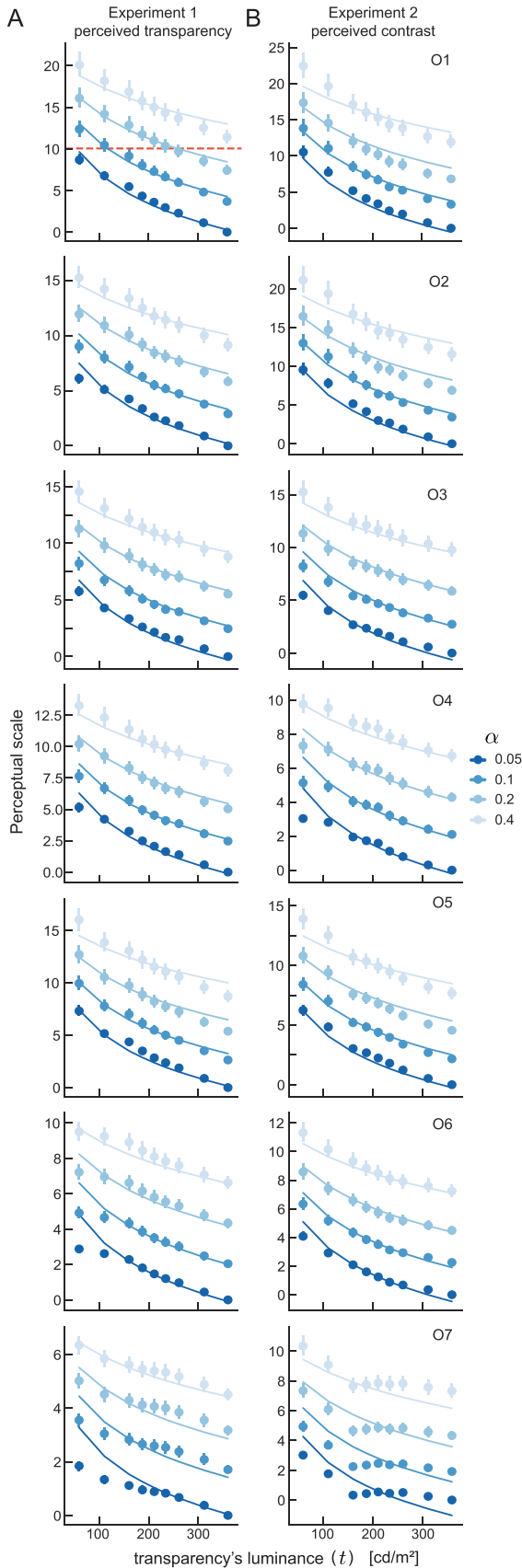


Figure 4. Perceptual scales obtained by MLCM for perceived transparency (A) and perceived contrast (B) for all observers (O1 to O7, rows). Error bars indicate 95% confidence intervals.

We evaluated six contrast metrics compiled by Moulden et al. (1990) and employed by Robilotto et al. (2002) (see Discussion). In the following definitions,  $l_i$  are the luminance values in the region of transparency and  $\bar{l}$  their arithmetic mean;  $n$  is the number of luminance values, which in our stimuli was 13.

- (i) Root mean square (RMS) contrast is defined as the standard deviation ( $SD$ ) of luminance values with respect to the global mean, and it is often used to predict perceived contrast (Peli, 1990).

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n (l_i - \bar{l})^2}$$

- (ii) Root mean square of log luminances, defined in the same way as (i) except that the standard deviation is computed for the logarithm of the luminances.

$$SDLG = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(l_i) - \overline{\log(l)})^2}$$

The following metrics are space averages, that is, the calculation is done for all possible nonequal combinations of luminances values ( $\forall i \neq j$ ).

- (iii) Space-averaged Michelson contrast

$$SAM = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left| \frac{l_i - l_j}{l_i + l_j} \right|$$

- (iv) Space-averaged logarithm of Michelson contrast

$$SAMLG = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \log \left| \frac{l_i - l_j}{l_i + l_j} \right|$$

- (v) Space-averaged Whittle contrast

$$SAW = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left| \frac{l_i - l_j}{\min(l_i, l_j)} \right|$$

← Continuous lines depict the prediction from the space-averaged logarithm of the Michelson contrast ( $SAMLG$ ). Dashed red line in O1 illustrates stimuli that would be perceived equally transparent.

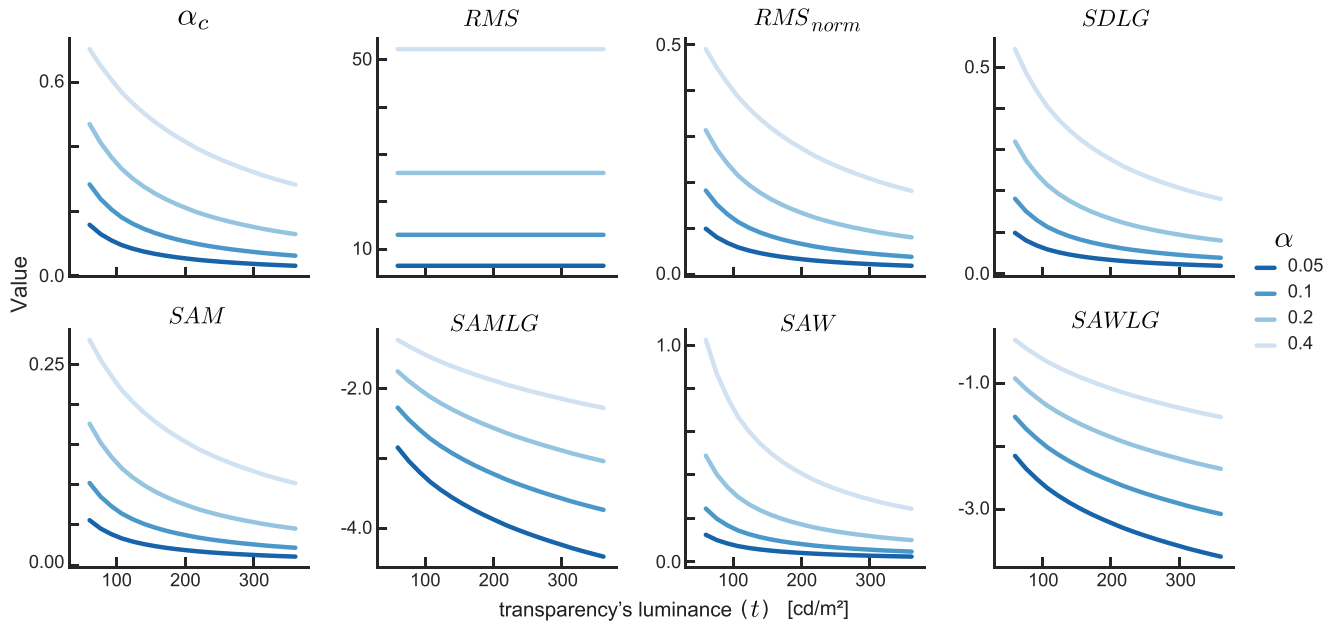


Figure 5. Contrast metrics and their dependency on the transparency's luminance (x-axis) and physical transmittance ( $\alpha$ ). Data plotted on the same format as Figure 4. Each panel shares the x-axis but not the y-axis, as each metric has a different range. See text for metrics formulae.

(vi) Space-averaged logarithm of Whittle contrast

$$SAWLG = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \log \left| \frac{l_i - l_j}{\min(l_i, l_j)} \right|$$

In addition to these six metrics, we also included

- (vii) An alternative formulation of RMS contrast in which the response is normalized with respect to the mean luminance (as defined by, e.g., Bex et al., 2009; Pelli & Bex, 2013)

$$RMS_{norm} = \frac{RMS}{\hat{l}} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (l_i - \bar{l})^2}}{\bar{l}}$$

- (viii)  $\alpha_c$  as defined in Equation 3.

$\alpha_c$  is based on Michelson contrast and hence depends only on the minimum and maximum luminance in the regions of transparency and plain view. The other metrics capture the entire distribution of luminance values in transparency. Note, however, that all metrics assume a preceding segmentation of the input.

Figure 5 shows the contrast metrics evaluated for our stimuli. The plots are analogous in design to those for the perceptual scales (Figure 4) in order to facilitate the comparison. By design, RMS contrast was constant across the mean luminance of the transparent medium.

This is inconsistent with the obtained scales and hence no adequate model of perceived transparency. All other metrics produce a similar pattern of results consistent with a trade-off between physical transmittance ( $\alpha$ ) and luminance  $t$ . One group of models, including  $RMS_{norm}$ ,  $\alpha_c$ ,  $SDLG$ ,  $SAM$ , and  $SAW$ , predicts an interaction effect between  $\alpha$  and  $t$  (Figure 5). The remaining two metrics,  $SAMLG$  and  $SAWLG$ , which compute the logarithm of contrast values, predict constant differences of perceived transparency as a function of  $t$  for different  $\alpha$ s. The  $SAMLG$  and  $SAWLG$  predictions are closer to the empirical scales.

To quantitatively test the goodness of fit for each of the tested metrics, we rescaled the prediction from each metric to each observer's individual response range. Different observers have different scale maxima, because they reflect each person's decision noise (the maximum value of an MLCM scale is inversely related to the amount of decision noise; see Figure 4 for interobserver variability in scale maxima).

We applied a linear transformation to rescale the prediction of each contrast metric to each observer's individual range. Minimizing the sum of squared errors (SSE), we used a single factor and intercept to rescale the four perceptual scales for different  $\alpha$ s. We did this separately for each contrast metric, observer, and experiment. Linear transformations are valid, because MLCM provides interval scales that do not change their numerical representation by linear transformations (Krantz et al., 1971; Gescheider, 1997).  $SAMLG$  was

Contrast metric	Exp.	SSE		<i>r</i> avg.
		Avg.	Range	
$\alpha_c$	1	47.68	[20.05, 90.74]	0.94
	2	47.87	[27.52, 82.68]	0.95
RMS	1	130.98	[20.48, 286.63]	0.86
	2	163.76	[39.23, 388.62]	0.86
RMS <sub>norm</sub>	1	55.71	[22.34, 105.85]	0.93
	2	56.40	[32.96, 96.90]	0.94
SDLG	1	67.75	[25.65, 128.25]	0.92
	2	69.44	[41.17, 118.54]	0.93
SAM	1	57.60	[22.87, 109.30]	0.93
	2	58.41	[34.26, 100.15]	0.94
SAMLG	1	8.73	[5.43, 13.48]	0.99
	2	15.98	[3.87, 34.50]	0.98
SAW	1	111.11	[36.77, 210.21]	0.86
	2	118.33	[70.75, 201.63]	0.88
SAWLG	1	7.15	[3.31, 10.48]	0.99
	2	12.34	[2.51, 25.28]	0.99

Table 1. Comparison between contrast metrics predictions and perceptual scales for both experiments. The sum of squared errors (SSE) and the Pearson's correlation coefficient (*r*) were calculated between predictions and scales for both experiments ("exp"). Mean ("avg") and range across observers.

one of the best-fitting metrics, and its predictions are shown as solid lines in Figure 4. Table 1 summarizes the goodness-of-fit evaluations for all contrast metrics. *SAMLG* and *SAWLG* produced the lowest sum of squared errors for all observers and in both experiments. The average Pearson's correlation coefficient between scales and metric prediction across observers was highest for *SAMLG* and *SAWLG* ( $r = 0.99$ ). Despite these relatively high average correlation coefficients, there was some variability across observers. The best fits with *SAMLG* were observed for O3 ( $r = 0.99$ , Experiment 1) and O6 ( $r = 0.99$ , Experiment 2), while the worse fit with *SAMLG* was observed for O7 in both experiments ( $r = 0.97$ ). The pattern was similar for *SAWLG*.

## Discussion

We studied whether perceived transparency can be predicted from image contrast. We used stimuli in which the transparent overlays covered multiple surface reflectances and contained geometric cues to depth such as occlusion and linear perspective (variegated checkerboards, Experiment 1). We used MLCM with paired comparisons to measure scales for perceived transparency. We also measured scales for perceived contrast for cut-out versions of the transparent region

of the same stimuli that did not contain cues to depth or transparency (Experiment 2).

Perceptual scales for both perceived transparency and perceived contrast were consistent across observers. Scales for perceived contrast closely resembled the scales for perceived transparency. We tested various contrast metrics to account for the perceptual scales and found that the space-averaged logarithm of contrasts, Michelson or Whittle contrast, accounted best for the data.

## Relation to previous work

In line with the pattern of matching results by Robilotto et al. (2002), the patterns of perceptual scales for perceived transparency and perceived contrast closely resembled each other. This similarity suggests that the two perceptual variables are related, and it is consistent with the notion of a common underlying mechanism. Different from Robilotto et al. (2002), our observed perceptual scales for both transparency and contrast were consistent with predictions by a space-averaged contrast metric relying either on log Michelson (*SAMLG*) or log Whittle contrast (*SAWLG*).

The studies were different in a number of ways, and we will discuss which of those differences might account for the different results. We used variegated checkerboards with geometric cues to transparency, whereas Robilotto et al. (2002) used coplanar "dead leaves" stimuli where the transparent region was moving. Both types of stimuli elicit vivid percepts of transparency, and hence we rule out stimulus differences as an explanation for the different results.

In the following, we discuss two factors that we consider most likely to account for the different results: (1) differences in calculating the contrast metrics and (2) differences in the experimental method (i.e., difference scaling vs. matching).

## Contrast metrics

We noticed a discrepancy in the definition of *SAMLG* and *SAWLG* between Robilotto et al. (2002) and Moulden et al. (1990). Robilotto et al. (2002) refer to Moulden et al. (1990) as the source for the contrast metrics, but Moulden et al. (1990) calculate *SAMLG* as  $\log \left| \frac{l_i - l_j}{l_i + l_j} \right|$ , whereas Robilotto et al. (2002) calculate it as  $\left| \frac{\log(l_i) - \log(l_j)}{\log(l_i) + \log(l_j)} \right|$ . The two formulations are not equivalent. The same discrepancy occurred for *SAWLG*. Using the original formulation by Moulden et al. (1990), *SAMLG* and *SAWLG* both predicted our data.

Both metrics rely on the logarithm of contrasts (Michelson or Whittle). Compressive nonlinearities—



such as the logarithm—have been applied to physical contrast in order to predict perceived contrast (e.g., Cannon & Fullenkamp, 1988, 1991) and it is also included in models of perceived contrast as the so-called nonlinear transducer function (e.g., Legge & Foley, 1980; Haun & Peli, 2013). Moulden et al. (1990) and Kingdom and Moulden (1991) argue that a contrast metric including a log transform is in agreement with physiological data that show that retinal ganglion cell responses are proportional not to their input (contrast) but to values transformed via a compressive nonlinearity. Kingdom and Moulden (1991) showed that a logarithmic contrast metric simultaneously accounted for brightness discrimination thresholds of increment and decrement pairs in data from Whittle (1986).

Kane and Bertalmio (2019) compared Kingdom and Moulden's (1991) model with three other models of brightness and contrast perception (Whittle's original formulation, a divisive gain model, and their own model) with respect to their ability to predict data from brightness detection, discrimination, and appearance experiments (Whittle, 1986, 1992). They showed that all models adequately predicted the data and that all models consist of conceptually analogous model components, namely, a monotonically increasing function and a contrast term. The log contrast term in Kingdom and Moulden (1991) was required to account for the asymmetry in discrimination performance between increments and decrements in Whittle's data, and here it is required to account for the dependency of perceived transparency and perceived contrast on the mean luminance of the transparent medium. The present results add to the list of effects that can be accounted for by such a contrast metric.

The log transform is only one of many possibilities of a compressive nonlinearity that could be used. Commonly, the Naka–Rushton function is used as a nonlinear transducer function in contrast models (e.g., Legge & Foley, 1980; Schütt & Wichmann, 2017) and can be parametrized to be identical to the log transform. Thus, a Naka–Rushton nonlinearity could also fit our data as well as the related Michaelis–Menten function,<sup>2</sup> which has recently been used to predict color scales (Knoblauch et al., 2020). The benefit of parametrized models is that the parameters might have a mechanistic interpretation. For example, Knoblauch et al. (2020) interpreted one of the parameters as contrast gain, and as such, it could be related to putative gain control mechanisms in spatial filtering models of the visual system. However, these additional parameters need to be fitted and sufficiently constrained by the data, and in order to connect model parameters at different levels of abstraction (physiological mechanism vs. computational model), additional “linking assumptions” are required. Here, we merely show that a log transform of a particular form seems to be capturing the data.

A potential limitation of the log contrast metrics occurs for cases where the transparent medium has low or zero reflectance (Kingdom, 2011). For a transparency of zero reflectance, *SAMLG* and *SAWLG* compute a value that is independent of the transparency's transmittance (Supplementary Fig. S4). However, Kingdom (2011) demonstrated (we informally confirm) that this does not correlate with perception. A zero-reflectance transparency rendered with higher transmittance looks indeed more transparent than one rendered with lower transmittance. This point was brought up by a reviewer, and unfortunately in the present study, we did not test low values of  $t$ . Thus, the explanatory power of these contrast metrics for low  $t$  remains an open question.

## Experimental method

Most previous studies on perceived transparency used asymmetric matching procedures (e.g., Singh & Anderson, 2002; Robilotto et al., 2002). Even though matching tasks are the “industry standard,” they are not ideal. In particular, when the matching is asymmetric (i.e., test and match stimulus are presented in different viewing contexts), observers might minimize perceived differences rather than produce perceived equality (e.g., Brainard et al., 1997; Ekroll et al., 2004; Logvinenko & Maloney, 2006). A difference scaling experiment avoids that issue, because observers judge the magnitude of differences instead of producing equality (when it might be impossible). In the following, we describe the putative relationship between perceptual scales and matches and the potential benefits of perceptual scales.

Figure 6 illustrates a matching experiment for the present scenario of measuring perceived transparency. Observers are presented with a standard stimulus of given transmittance  $\alpha_{st}$  and luminance  $t_{st}$  of the transparent medium. The transmittance of the test stimulus  $\alpha_{test}$  is also given, and observers adjust the luminance of the transparent medium  $t_{test}$  such that it looks equally transparent to that of the standard (Figure 6A). The output of this procedure is a perceptual match quantified in physical units, namely, the luminances of the test and standard transparent media (Figure 6C). The procedure does not reveal the function of interest, which is the perceptual scale (or transducer function; Kingdom & Prins, 2010), which relates the transparency's luminance ( $t$ ) to its perceived transparency (Figure 6B).

Instead, matching procedures make an (implicit) linking assumption about how matches relate to the internal scale (e.g., Maertens & Shapley, 2013; Teller, 1984; Brindley, 1960). Figure 6B depicts a number of hypothetical internal scales relating the luminance of a transparent medium,  $t$ , to its perceived transmittance

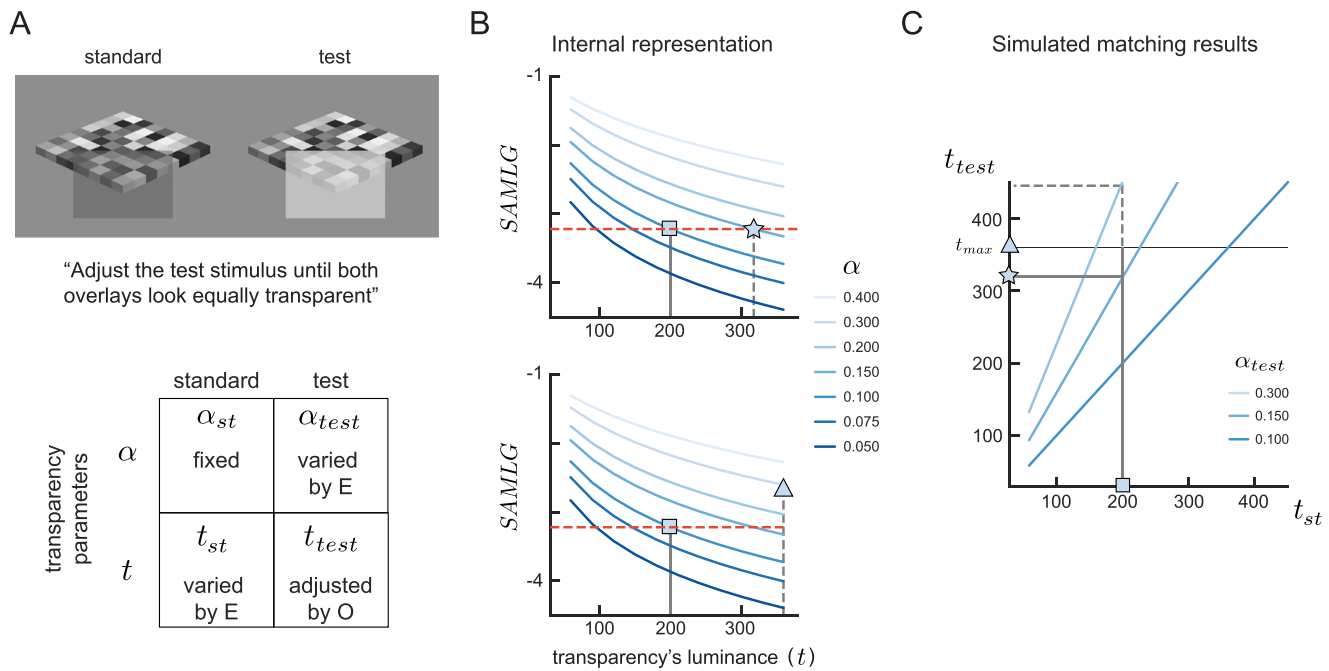


Figure 6. Linking assumption between (asymmetric) matching and perceptual scales. (A) Experimental design of an asymmetric matching experiment. The transparent medium of the standard stimulus has one fixed transmittance ( $\alpha_{st}$ ) and a fixed number of luminances ( $t_{st}$ ). The observer adjusts the luminance of the transparency of the test stimulus ( $t_{test}$ ) for different levels of transmittance ( $\alpha_{test}$ ) of the test stimulus to match the perceived transparency of the test to that of the standard. (B) Internal scales relating luminance of the transparency (x-axis) and transmittance (panel variable) to perceived transmittance (y-axis). Perceived transmittance is computed as the space-average logarithm of Michelson contrast (SAMLG). The top panel shows a case where a perceptual match can be obtained between standard (square) and test (star). The bottom panel shows a case where only a partial match can be obtained (triangle; see text for explanation). (C) Data in a hypothetical matching experiment. The lines are derived as predictions from the scales in B. The dashed line indicates where the match for that  $\alpha_{test}$  should be if a test stimulus with that luminance would have been included. Since that luminance is beyond the range of possible test luminances, the observer sets the test luminance to  $t_{max}$ , the maximum luminance (triangle).

as a function of different physical transmittances,  $\alpha$ . Consider first the upper panel. The standard stimulus, indicated by the square, has a fixed transmittance of  $\alpha = 0.1$  and a luminance of  $l = 200 \frac{cd}{m^2}$ . The test stimulus for that trial has a transmittance of  $\alpha = 0.15$ . At the start of a matching trial, the luminance of the test is set to a random value. To find a perceptual match, the scale value of the standard is extrapolated (dashed red line) to the scale that corresponds to the  $\alpha$  of the test stimulus (star symbol). The match luminance of the test is found by reading out the x-value at that test scale value ( $\sim 320 \frac{cd}{m^2}$ ).

The lower panel in Figure 6B depicts the same scenario but for a test stimulus with transmittance  $\alpha = 0.3$ . Following the same routine as described above (i.e., extrapolating the scale value for the standard in order to find the intersection with the scale of the test), it is evident that there is no intersection point across the range of tested luminances. In other words, for an  $\alpha$  of 0.3, the luminance of the transparency would have to be very light to look equally transparent as

that of the standard. This would be a case in which an observer might do an as-close-as-possible but not an actual match. Figure 6C shows the data points for such a hypothetical matching experiment. The solid lines indicate matches predicted from the scales in Figure 6B. The star and square symbols show data from one matching trial, where the matching was possible ( $\alpha = .15$ ). The triangle and the square symbols demarcate a data point that lies beside the predicted light blue matching function, because a good match was not possible. This illustrates that matching data are only an indirect way to estimate perceptual representations of physical variables and that they might not be a good estimate when matching is not possible.

Scaling procedures such as MLCM provide an empirical estimate for the function that relates the stimulus and the perceptual dimension of interest. Difference scaling also avoids the problem of asking for perceptual equality. Hence, we think that perceptual scales complement and in some ways extend the data from standard matching routines.

## Perceived contrast and perceived transparency

Robilotto et al. (2002) suggested that “perceived image contrast is the sensory determinant of perceived transparency” because they observed similar reflectivity and inner transmittance matches for these two perceptual variables. Our data show that perceived transparency as well as perceived contrast can be predicted by a measure that relies on physical image contrast. This seems to point at least to a common mechanism that determines both perceived contrast and perceived transparency.

However, we cannot draw further conclusions about the relationship between perceived contrast and transparency with the present data. Scales for perceived transparency were measured in variegated checkerboards, whereas those for perceived contrast were measured in cut-outs. The scales from the two experiments are independent from each other, and the scales within each experiment are anchored relative to one another. By design, MLCM anchors the scale with the smallest value to zero and the scale with the highest value to a maximum value, which represents the inverse of the estimated judgment noise. Thus, the individual scales from the two experiments cannot be directly compared as they do not share a common metric (see Aguilar & Maertens, 2020, for a more detailed treatment of the issue of comparing scales across different contexts).

A different experiment would be required to determine the empirical relationship between perceived contrast and transparency. In such an experiment, observers would have to judge perceived contrast in full variegated checkerboards, in cut-outs and in mixed comparisons of both types of stimuli. There would be three dimensions to manipulate (transmittance, luminance of the transparent medium when opaque, and stimulus version—full or cut-out). This would result in more than 25,000 trials in total per observer (with stimuli otherwise comparable to the current design). The rapid increase in number of trials for this type of scaling procedure (Maximum Likelihood Difference Scaling (MLDS), MLCM) renders the procedures less attractive from a pragmatic point of view and might sometimes outweigh the theoretical benefits discussed above. Alternative strategies such as subsampling (Knoblauch & Maloney, 2012; Abbatecola et al., 2021), having a reduced number of stimuli per dimension (Sun et al., 2021), or the use of so-called embedded methods from the machine learning community (see Haghiri et al., 2020, for example) are currently being explored and might allow more efficient ways of perceptual scale measurements in the future.

## Open questions

We mentioned above that all contrast metrics assume a (magical) segmentation step that assigns

which luminance values belong to the transparent overlay and which belong to regions in “plain view.” In other words, the tested contrast metrics are not image-computable. The metric  $\alpha_c$  takes as input the values from both regions (Equation 3), whereas all other metrics considered only the luminance values belonging to the transparent region. Thus, in the present study, the segmentation assumption is not crucial because we could calculate contrast metrics (except  $\alpha_c$ ) based on the luminance values from the entire stimulus. The pattern of predictions would be the same because the luminance values in “plain view” were identical in all stimuli. Including this region in the computation of the contrast metrics would only introduce a constant additive shift without changing the overall pattern seen in Figure 5.

However, this approach does not generalize to other types of stimulus manipulations (i.e., variation of luminance range and contrast in the region of plain view). Changing the contrast surrounding a target region affects its perceived contrast (e.g., the contrast-contrast phenomenon; Chubb et al., 1989), as well as its perceived transparency (Robilotto & Zaidi, 2004). These “induction” effects cannot be captured by the type of contrast metrics studied here, as they presuppose the segmentation of the image into different regions, and they do not consider spatial relationships between different regions.

A possible direction of further research would be to compare perceptual scales against more sophisticated models of perceived contrast that include spatial relationships of some kind. For example, Haun and Peli (2013) have presented a multiscale spatial filtering model that aims to predict perceived contrast in natural images. However, relating the output of such a model (which is in itself an image) to our perceptual scales would require the formulation of additional linking assumptions. This is a research endeavor in itself and clearly beyond the scope of the present study.

## Conclusions

We used MLCM to measure perceptual scales of perceived transparency in images of variegated checkerboards with multiple reflectances and cues to transparency. We also measured perceptual scales for perceived contrast in cut-out versions of the checkerboards that did not evoke a transparency percept. We found that both judgments were similar to each other, suggesting a common mechanism. Taking advantage of the type of data obtained by the MLCM method, we could determine that both judgments could be accounted for by a metric relying on the logarithm of Michelson or Whittle contrast. These contrast metrics can be readily calculated from the array of luminance values known to belong to the transparency region.

*Keywords:* perceived transparency, perceived contrast, MLCM, conjoint measurement

## Acknowledgments

The authors thank two anonymous reviewers for their constructive suggestions that helped improve this manuscript.

Supported by a research grant of the German Research Foundation (DFG MA5127/5-1).

Commercial relationships: none.

Corresponding author: Guillermo Aguilar.

Email: guillermo.aguilar@mail.tu-berlin.de.

Address: Computational Psychology, Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany.

## Footnotes

<sup>1</sup>We use the term “perceived transparency” interchangeably with “perceived transmittance” to describe the subjective impression of how much is seen through a transparent medium.

<sup>2</sup>The Naka–Rushton function reduces to a Michaelis–Menten function when the exponents are equal to 1.

## References

- Abbatecola, C., Gerardin, P., Beneyton, K., Kennedy, H., & Knoblauch, K. (2021). The role of unimodal feedback pathways in gender perception during activation of voice and face areas. *Frontiers in Systems Neuroscience*, *15*, 669256, doi:10.3389/fnsys.2021.669256.
- Aguilar, G., & Maertens, M. (2020). Toward reliable measurements of perceptual scales in multiple contexts. *Journal of Vision*, *20*(4), 19, doi:10.1167/jov.20.4.19.
- Anderson, B. L. (2008). Transparency and occlusion. In R. H. Masland et al. (Eds.), *The senses: A comprehensive reference* (Vol. 2, pp. 239–244). New York: Academic Press, doi:10.1016/B978-012370880-9.00319-4.
- Anderson, B. L., Singh, M., & Meng, J. (2006). The perceived transmittance of inhomogeneous surfaces and media. *Vision Research*, *46*(12), 1982–1995, doi:10.1016/j.visres.2005.11.024.
- Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision*, *9*(10), 1–19, doi:10.1167/9.10.1.
- Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image. 1. asymmetric matches. *Journal of the Optical Society of America A*, *14*(9), 2091–2110, doi:10.1364/JOSAA.14.002091.
- Brindley, G. (1960). *Physiology of the retina and visual pathway*. Baltimore, MD: William & Wilkins.
- Cannon, M. W., & Fullenkamp, S. C. (1988). Perceived contrast and stimulus size: Experiment and simulation. *Vision Research*, *28*(6), 695–709, doi:10.1016/0042-6989(88)90049-1.
- Cannon, M. W., & Fullenkamp, S. C. (1991). A transducer model for contrast perception. *Vision Research*, *31*(6), 983–998, doi:10.1016/S0042-6989(05)80001-X.
- Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences*, *86*(23), 9631–9635, doi:10.1073/pnas.86.23.9631.
- Ekroll, V., Faul, F., & Niedere, R. (2004). The peculiar nature of simultaneous colour contrast in uniform surrounds. *Vision Research*, *44*(15), 1765–1786, doi:10.1016/j.visres.2004.02.009.
- Gescheider, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Haghiri, S., Wichmann, F. A., & von Luxburg, U. (2020). Estimation of perceptual scales using ordinal embedding. *Journal of Vision*, *20*(9), 14, doi:10.1167/jov.20.9.14.
- Haun, A. M., & Peli, E. (2013). Perceived contrast in complex images. *Journal of Vision*, *13*(13), 3, doi:10.1167/13.13.3.
- Kane, D., & Bertalmio, M. (2019). A reevaluation of Whittle (1986, 1992) reveals the link between detection thresholds, discrimination thresholds, and brightness perception. *Journal of Vision*, 1–13, doi:10.1167/19.1.16.
- Kasrai, R., & Kingdom, F. (2001). Precision, accuracy, and range of perceived achromatic transparency. *Journal of the Optical Society of America A*, *18*(1), 1, doi:10.1364/JOSAA.18.000001.
- Kingdom, F. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, *51*(7), 652–673, doi:10.1016/j.visres.2010.09.012.
- Kingdom, F., & Moulden, B. (1991). A model for contrast discrimination with incremental and decremental test patches. *Vision Research*, *31*(5), 851–858.
- Kingdom, F., & Prins, N. (2010). *Psychophysics: A practical introduction*. London, UK: Academic Press.
- Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R*. New York, NY: Springer.

- Knoblauch, K., & Maloney, L. T. (2014). Mlcm: Maximum likelihood conjoint measurement [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=MLCM>.
- Knoblauch, K., Marsh-Armstrong, B., & Werner, J. S. (2020). Suprathreshold contrast response in normal and anomalous trichromats. *Journal of the Optical Society of America A*, 37(4), A133, doi:10.1364/josaa.380088.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Vol. I. Additive and polynomial representations*. Mineola, NY: Dover Publications.
- Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America A*, 70(12), 1458–1471, doi:10.1364/JOSA.70.001458.
- Logvinenko, A. D., & Maloney, L. T. (2006). The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching. *Perception & Psychophysics*, 68(1), 76–83, doi:10.3758/BF03193657.
- Luce, R., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1(1), 1–27, doi:10.1016/0022-2496(64)90015-X.
- Maertens, M., & Shapley, R. (2013). Linking appearance to neural activity through the study of the perception of lightness in naturalistic contexts. *Visual Neuroscience*, 30(5–6), 289–298, doi:10.1017/S0952523813000229.
- Maloney, L. T., & Knoblauch, K. (2020). Measuring and modeling visual appearance. *Annual Review of Vision Science*, 6, 519–537, doi:10.1146/annurev-vision-030320-041152.
- Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, 13, 59–66.
- Metelli, F. (1974). Achromatic color conditions in the perception of transparency. In R. B. MacLeod, & H. L. Pick (Eds.), *Perception: Essays in honor of J. J. Gibson*. Ithaca, NY: Cornell University Press.
- Moulden, B., Kingdom, F., & Gatley, L. F. (1990). The standard deviation of luminance as a metric for contrast in random-dot images. *Perception*, 19(1), 79–101, doi:10.1068/p190079.
- Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America A*, 7(10), 2032, doi:10.1364/JOSAA.7.002032.
- Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision Research*, 90, 10–14, doi:10.1016/j.visres.2013.04.015.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org>.
- Robilotto, R., Khang, B.-G., & Zaidi, Q. (2002). Sensory and physical determinants of perceived achromatic transparency. *Journal of Vision*, 2(5), 388–403, doi:10.1167/2.5.3.
- Robilotto, R., & Zaidi, Q. (2004). Perceived transparency of neutral density filters across dissimilar backgrounds. *Journal of Vision*, 4(3), 5, doi:10.1167/4.3.5.
- Schütt, H. H., & Wichmann, F. A. (2017). An image-computable psychophysical spatial vision model. *Journal of Vision*, 17(12), 1–35, doi:10.1167/17.12.12.
- Singh, M., & Anderson, B. L. (2002). Toward a perceptual theory of transparency. *Psychological Review*, 109(3), 492–519, doi:10.1037/0033-295X.109.3.492.
- Singh, M., & Anderson, B. L. (2006). Photometric determinants of perceived transparency. *Vision Research*, 46(6–7), 879–894, doi:10.1016/j.visres.2005.10.022.
- Sun, H.-C., St-Amand, D., Baker, C. L., & Kingdom, F. (2021). Visual perception of texture regularity: Conjoint measurements and a wavelet response-distribution model. *PLoS Computational Biology*, 17(10), e1008802, doi:10.1371/journal.pcbi.1008802.
- Teller, D. Y. (1984). Linking propositions. *Vision Research*, 24(10), 1233–1246, doi:10.1016/0042-6989(84)90178-0.
- Tudor-Hart, B. (1928). Studies in transparency, form and colour. *Psychologische Forschung*, 10(1), 255–298, doi:10.1007/BF00492012.
- Whittle, P. (1986). Increments and decrements: Luminance discrimination. *Vision Research*, 26(10), 1677–1691, doi:10.1016/0042-6989(86)90055-6.
- Whittle, P. (1992). Brightness, discriminability and the “Crispening Effect.” *Vision Research*, 32(8), 1493–1507, doi:10.1016/0042-6989(92)90205-W.
- Wiebel, C., Aguilar, G., & Maertens, M. (2017). Maximum likelihood difference scales represent perceptual magnitudes and predict appearance matches. *Journal of Vision*, 17(4), 1–14, doi:10.1167/17.4.1.