Delft University of Technology

Perception of object illumination depends on highlights and shadows, not shading

te Pas, Susan; Pont, Sylvia; Dalmaijer, ES; Hooge, ITC

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Perception of object illumination depends on highlights and shadows, not shading

Susan F. te Pas

Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, the Netherlands

Sylvia C. Pont

Perceptual Intelligence Lab, Faculty of Industrial Design Engineering, Delft University of Technology, Delft, the Netherlands

Edwin S. Dalmaijer

Department of Experimental Psychology, Medical Sciences Division, University of Oxford, Oxford, United Kingdom

Ignace T. C. Hooge

Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, the Netherlands

Human observers are able to successfully infer direction and intensity of light from photographed scenes despite complex interactions between light, shape, and material. We investigate how well they are able to distinguish other low-level aspects of illumination, such as the diffuseness and the number of light sources. We use photographs of a teapot, an orange, and a tennis ball from the ALOI database (Geusebroek, Burghouts, & Smeulders, 2005) to create different illumination conditions, varying either in diffuseness of a single light source or in separation angle between two distinct light sources. Our observers were presented with all three objects; they indicated which object was illuminated differently from the other two. We record discrimination performance, reaction times, and eye fixations. We compare the data to a model that uses differences in image structure in same-object comparisons, and outcomes suggest that participants mostly rely on the information contained in cast shadows and highlights. The pattern of eye fixations confirms this, showing that after the first fixation, observers mostly fixate cast shadow areas. However, information in the highlights is rather salient, so it might be available from first fixation, making separate fixations are unnecessary.

Introduction

Appearance is the result of a physical projection of an object or scene on the retina, combined with our brain's perceptual and cognitive interpretation of this projection. We can describe the physical appearance of an object as a combination of the light field, its material properties, and its shape. In return, object appearance also provides us with clues about the illumination, shape, and material properties. However, deriving the illumination, shape, and material properties from images or even from real scenes is by no means a trivial task. Mathematically, there usually is no unique solution to the problem; many combinations of objects and light fields would result in the same images and retinal input. Calculating shape, illumination, and material from images, be it on the retina or in photographs, is therefore an underdetermined problem (Adelson & Pentland, 1996; Belhumeur, Kriegman, & Yuille, 1999; Blake & Bülthoff, 1990; Dror, Willsky, & Adelson, 2004).

In many cases, it is the shape or material of objects and scenes that we humans are interested in and not the illumination per se. Generally, observers seem to judge material and shape perceptually without giving the ambiguities that are always present any thought. To determine the perceived shape or material, however, observers need to disentangle the effect of illumination on the appearance of objects from the effects caused by shape and material properties. To be able to do that means that assumptions about the light field have to be made.

A lot of work on the separation of perceived illumination and material has been conducted in the field of color constancy and the perception of lightness and brightness. Typically, a few rendered two-dimen-

sional surfaces are used in so-called Mondrian stimuli that show that illumination and material are actually metameric. Although many results have been achieved in this manner, questions have been raised as to whether these results can actually be extrapolated to more complex or even natural scenes (Gilchrist, 1994; Hurlbert, 1999; Kraft & Brainard, 1999; Radonjić et al., 2016; te Pas & Koenderink, 2004). In the Mondrian case, the relevant information about the illuminant is mainly in the spectral distribution of the light, not the spatial distribution. However, there has been a growing interest in investigating interactions that emerge in complex, typically three-dimensional, scenes (e.g., Boyaci, Fang, Murray, & Kersten, 2007; Braje, Legge, & Kersten, 2000; Kartashova, de Ridder, te Pas, Schoemaker, & Pont, 2015; Kartashova, Sekulovski, de Ridder, te Pas, & Pont, 2016; Ling & Hurlbert, 2004; Marlow, Kim, & Anderson, 2012; Nishida & Shinya, 1998; Obein, Knoblauch, & Vienot, 2004; Pont & Koenderink, 2007; Robilotto & Zaidi, 2004; Xia, Pont, & Heyndericx, 2014). The context provided in the scene can help in the disambiguation between light and material, but even in complex, photorealistic images, material changes are often confounded with illumination changes (Anderson, 2011; Boyaci, Maloney, & Hersh, 2003; Doerschner, Boyaci, & Maloney, 2007; Fleming, 2014; Fleming, Dror, & Adelson, 2003; Maloney, Gerhard, Boyaci, & Doerschner, 2010; Pont, Koenderink, Van Doorn, Wijntjes, & te Pas, 2012; Pont & Te Pas, 2006; Ripamonti et al., 2004; te Pas & Pont, 2005; Toscani, Zdravković, & Gegenfurtner, 2016; Zhang, de Ridder, & Pont, 2015). However, when asked to place photographed real materials in simple natural categories, participants are surprisingly fast, even when stimuli are degraded and materials are taken out of context (Sharan, Rosenholtz & Adelson, 2009, 2014; Wiebel, Valsecchi, & Gegenfurtner, 2013).

When judging the illumination in a scene, observers rely on several cues. Most of the cues that are used have to do with the way objects reflect the light and thus with object properties and scene layout. They enable the human observer to determine certain lower-order properties of the light field. Texture shading and cast shadows can, for instance, be used to determine the direction and intensity of a light source (Casati, 2004; Koenderink, van Doorn, Kappers, te Pas, & Pont, 2003; Koenderink, van Doorn, & Pont, 2004; Xia et al., 2014). Morgenstern, Geisler, and Murray (2014) determined that observers are attuned to diffuseness levels of natural scenes. Observers even can infer low-level properties of the light in empty space (Koenderink, Pont, van Doorn, Kappers, & Todd, 2007; Schirillo, 2013) from complex photographed real scenes and also from actual scenes (not photographs) by using a light probe (Kartashova et al., 2016; Koenderink et al., 2007; Xia et al., 2014).

However, human observers are not very good at spotting local inconsistencies in the lighting or at using cues such as shading and shadowing when there are very few shape cues (O'Shea, Agrawala, & Banks, 2010; Ostrovsky, Cavanagh, & Sinha, 2005). Moreover, they can distinguish only a limited number of cases of global light field structures (Kartashova et al., 2015; Kartashova et al., 2016; van Doorn, Koenderink, Todd, & Wagemans, 2012). Appearance of gloss, for instance, is best mitigated by simply shaped highlights, which are usually driven by simple light fields (van Assen, Wijntjes, & Pont, 2016). Perceptually, observers can typically distinguish differences in contrast gloss and in distinction of image gloss; both types depend on the interaction of light with the object and on the surrounding scene (Ferwerda, Pellacini, & Greenberg, 2001). Although gloss is rather complex, observers are typically very fast at recognizing glossy (and other) materials (Sharan et al., 2009, 2014). When cues are not conclusive, observers make assumptions about the light field that are probably based on prior experience, such as that light comes from above (Fleming, 2012; Mamassian, 2004; Mamassian & Goutcher, 2001; Morgenstern, Murray, & Harris, 2011).

Apparently, when enough cues such as cast shadows, shading, and highlights are available, and the context is rich enough, human observers are rather good at determining basic aspects such as the general direction and intensity of the light. More complex aspects of illumination are often confused with changes in material properties or shape, even in rich contexts. The question we ask here is how well human observers are able to distinguish basic aspects of the illumination that can generate similar but not identical light fields in space, such as the variations of the diffuseness of a single light source and the separation distance between multiple light sources, when we offer a sparse but well-controlled context. Moreover, we would like to know what kind of information observers use to assess these qualities of the light field by also recording eye fixations and reaction times during their assessment of the scene. Boyaci, Doerschner, and Maloney (2006) have shown that highlights, shadows, and shading information are all used in determining lighting conditions in complex scenes. To determine if that is also true in our rather sparse contexts, we ask observers to determine which one of three objects is illuminated differently from the other two. We measure percentage correct, reaction times, and eye fixations on the objects. We compare results of human observers with models that use different types of information about the illumination, specifically information contained in the highlights, shading, or shadow of the object at hand.

Figure 1. Photographs of the teapot (No. 161 from the ALOI database; Geusebroek et al., 2005; illumination conditions l1c1–l5c1 rendered in black and white). These five photographs were used to render the different illumination combinations used in the experiment.

Methods

Stimuli

We used pictures of a teapot, an orange and, a tennis ball from the Amsterdam Library of Object Images database (ALOI; Geusebroek, Burghouts, & Smeulders, 2005) to create six illumination conditions for each object. From the ALOI database we, selected pictures of each object shot with a fronto-parallel camera position, illuminated with a single light source in five different positions: straight above, 15° and 30° to the right and 15° and 30° to the left (for an example, see Figure 1).

Unfortunately, the illumination conditions in this database do not vary in diffuseness or in separation of the sources. However, because they are so well controlled, we can create new pictures with combined illuminations by a simple superposition of the separate images. Because light follows the superposition principle, the result is physically correct. This can be done with different weights to create a whole range of different illumination conditions.

We combined these five pictures by superposing them with different weights to create new illumination conditions that resemble either a single source with different diffuseness or two distinct light sources with different separation. To determine the weights, we implemented a Gaussian weighting function with its top (largest weight) in the center and varied the diffuseness (spread) by varying the width of the Gaussian weight function so that at standard deviation, the width is 10°, 20°, or 30° to create the impression of light sources with different diffuseness. By defining as a weight function two Gaussians with constant width of 10°, but with their peaks at a distance from the center, we were able create the appearance of two distinct light source directions with a varying separation angle of 10°, 20°, and 30° (see Figure 2 for a graphic representation of this weighting process).

The illumination conditions we created in this way are illumination from a single direction with three values for diffuseness (10°, 20°, and 30° of width of the illuminant) and illumination from two distinct directions with three amounts of separation between the two sources (10°, 20°, and 30° of angle; for the resulting stimuli, see Figure 3).

Participants

A total of 16 observers participated in the actual experiment (five men, 11 women). All participants were students of Utrecht University who were paid for their participation. All observers were naïve with respect to the purpose and nature of the experiment. All observers gave written informed consent. Experiments were done in agreement with local ethical guidelines, Dutch Law, and the Declaration of Helsinki.

Task

In the instruction phase, observers were shown pictures like the ones in the real experiment but of a different object and with large differences in illumination to illustrate the different types of illumination they could expect. They were given only examples; no instructions were given as to how they should look at or compare the images. In the experimental phase, observers were presented with a tennis ball, a teapot, and an orange side by side on a black background on every trial. They had to indicate which of the three objects was illuminated differently from the other two (an odd-one-out task). Participants reported that they found this a relatively easy task to do. There are 30 different illumination combinations (6 × 5; of the odd-one-out illumination and the illumination of the other two objects). For each of these combinations, we have three objects that can be the odd-one-out times six possible layouts of the three objects, yielding 18 different trials per condition. This means that in total, there are 540 trials per participant. The eye fixations of the participants were recorded during the entire experiment. Although we did not instruct participants to react as fast as possible, we did record the reaction time for each condition as well.
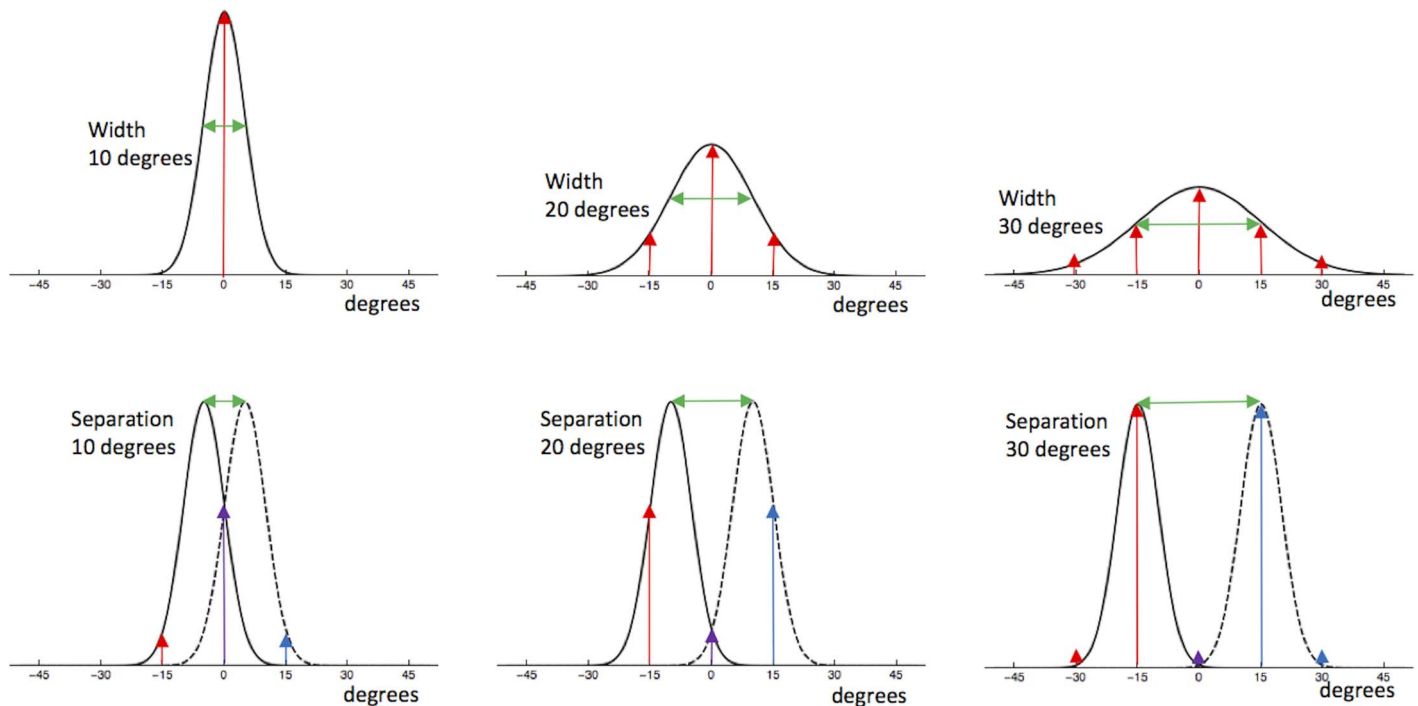
Figure 2. Graphic illustration of the weighting function for the addition of images. Top row: Relative image weights for diffuseness (for the construction of an image with a diffuse light source with a 10°, 20°, and 30° of angle width, respectively). Bottom row: Relative weights for separation (for the construction of an image with two light sources with a separation of 10°, 20°, and 30° of angle, respectively).

Setup

In the experimental phase, observers were seated with their head in a chinrest at 57 cm from a 22-in. LaCie Blue Electron CRT monitor. Experiments were performed using a PC running the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). Eye fixations were recorded at 52 Hz using a portable, EasyGaze™ eye tracker (Design Interactive, Inc, Oviedo, FL).
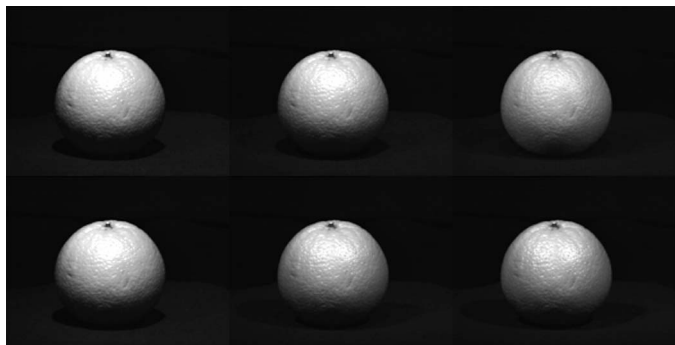

Figure 3. Photographs of the actual oranges stimuli we used. Top row: One light source with varying diffuseness (from left to right: 10°, 20°, and 30° spread). Bottom row: Two light sources with varying separation (from left to right: 10°, 20°, and 30° apart).

Analysis of data

For each illumination condition, we obtain an average percentage correct and an average reaction time per participant. The eye fixations of the participants were analyzed offline. Fixation detection was done using a custom Matlab program that marked fixations by an adaptive velocity threshold method (Hooge & Camps, 2013).

From the analyzed fixation locations, we create average heat plots using the output of the probability density function of the fixation locations, based on a smooth kernel density estimate using Silverman's rule to determine bandwidth (this is a standard Mathematica function) per participant for the overall fixation location and for the location of the first five fixations separately. We also compute the percentage of fixations in different areas of interest (AOI) within the stimulus, such as a highlight AOI, a shading AOI, and a cast shadow AOI. For the eye fixations, we kept the AIOs simple and coarse, so that they all span an equally sized area. We used the upper third of the scene to define the location of highlights, the middle third for the location of shading, and the lower third for the location of cast shadows. In this way, the AOIs are the same size; however, there is some overlap in the type of information contained in each AOI. From the heat maps, we can see that the fixations are mainly
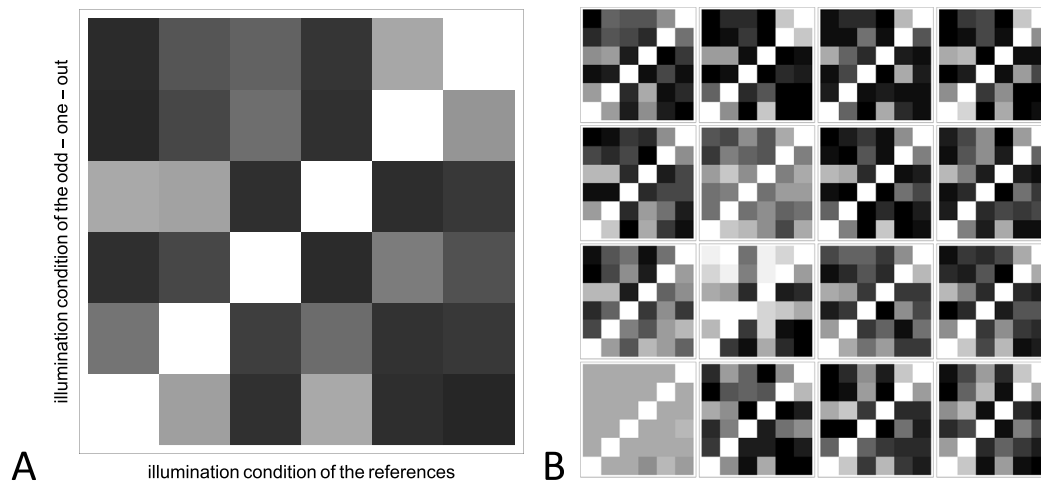
Figure 4. Percentages correct for all 30 combinations of illumination conditions. On the horizontal axis is the illumination condition of the two reference objects and on the vertical axis the illumination condition of the odd-one-out (left to right and bottom to top: first three increasing diffuseness of 10°, 20°, and 30° spread, respectively; second three increasing separation of 10°, 20°, and 30°, respectively). (A) Averaged over all 16 participants. (B) Individual data. Black means 100% correct; white means 0% correct.

concentrated in the areas we defined. For both modeling and behavioral analysis, we use somewhat more sophisticated AOIs that were defined by masks that were created by cutting out the relevant parts of the scene in terms of highlights, cast shadows, or shading by hand.

Results

Psychophysics

We collected both percentages correct and reaction times for all observers in all conditions. To be able to compare results, we plot these in matrix form. On the horizontal axis is the illumination condition of the two objects that have the same illumination (references) and on the vertical axis the illumination condition of the odd-one-out (test). We visualize percentages correct by means of a gray scale, with black meaning 100% correct and white meaning 0% correct (with 33% chance level). The conditions on the diagonal where the conditions of both reference and test were the same, expressed here in white, were not included in the experiment. This way of visualizing enables us to look at patterns of results and compare them. Figure 4 shows the average percentages correct over all observers as well as individual results. Although observers reported that this was a task they could perform, they still made a lot of errors. There is a distinct pattern of errors, in that some illumination conditions are clearly more difficult to distinguish than others. Specifically, although overall performance is low, the percentages correct increase when the difference in

diffuseness is larger (lower-left quadrant). However, there is no clear increase in performance when the separation difference is larger (upper-right quadrant). Two distinct light sources and one diffuse light source are frequently mixed up.

Although we did not ask them to respond as fast as possible, we also collected average reaction times per condition from our observers. To be able to compare the pattern of results from the average reaction times to the pattern of results of the percentages correct, we visualize the reaction times in the same way, with black meaning the shortest average reaction time and white meaning the longest average reaction time. Again, the same-illumination conditions on the diagonal were not included in the experiment. Figure 5 shows the average reaction times over all observers as well as individual results. Again, there is a distinct pattern, in that some illumination conditions clearly take more time to distinguish than others. Specifically, just like with the percentages correct, the reaction times decrease with larger diffuseness difference (lower-left quadrant). However, there is no clear decrease with decreasing separation (upper-right quadrant).

The pattern of results is slightly different from the pattern of results we found for the percentages correct; however, there is a strong negative correlation (Pearson correlation = 0.92) between these two measures: Shorter reaction times generally mean more correct judgments. This is visualized in Figure 6, where we plotted the average percentage correct for a particular illumination combination against the average reaction time for the same illumination combination.
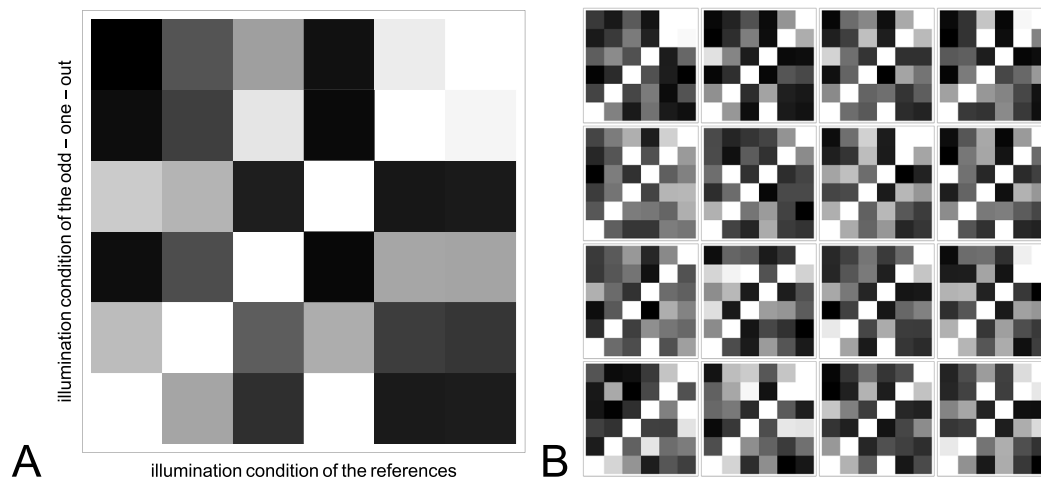
Figure 5. Reaction times for all 30 combinations of illumination conditions. (A) Averaged over all 16 participants. (B) Individual data. Illumination conditions are ordered in the same way as in Figure 4. Black means shortest reaction time; white means longest reaction time.

Modeling

To investigate what type of information participants were using to perform this task, we modeled performance of observers based on different assumptions.

In our first 4 models, termed PixelDifferencesModel, MeanLuminanceModel, BrightestLuminanceModel, and LuminanceSkewnessModel, we take some well-known image statistics and see whether these image statistics can predict the results of our participants. In the PixelDifferencesModel, we assume that observers simply based their judgments on absolute pixel differences between the three images. To model this, for each condition, we took the pixel values of the two images that were illuminated in the same way and subtracted the pixel values of the image with the odd-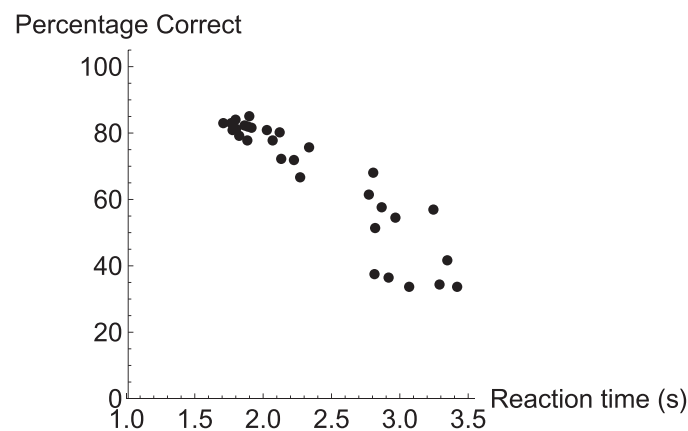one-out illumination from them. We averaged the resulting differences. If an observer were to use this strategy, one would obtain the pattern of results that is shown in Figure 7A, where black means highest pixel difference and white means lowest pixel differences between illumination conditions. In the MeanLuminanceModel, we simply took the average luminance of all three images and calculated which average luminance was most different from the other two. If an observer were to use this strategy, one would obtain a pattern of results shown in Figure 7B. A number of studies have reported that the luminance in the brightest part of the image or the skewness of the luminance histogram yields information on material properties such as gloss (Motoyoshi, Nishida, Sharan, & Adelson, 2007; Sharan, Li, Motoyoshi, Nishida, & Adelson, 2008; Toscani, Valsecchi, & Gegenfurtner, 2013, 2017; Wiebel, Toscani, & Gegenfurtner, 2015). These statistics might also be of interest for the perception of illumination, and so we modeled them too to see whether they can predict the results of our participants. For the BrightestLuminance model, we chose not to use the brightest pixel in the image but instead used the 95% quantile of the brightness histogram, because in photographs of natural images, the brightest pixel is usually very artificial because of things such as glare. If an observer were to use this strategy, one would obtain a pattern of results shown in Figure 7C. In the LuminanceSkewnessModel, we took the skewness of the luminance histogram of all three images and calculated which skewness luminance was most different from the other two. If an observer were to use this strategy, one would obtain a pattern of results shown in Figure 7D.

Pearson correlation coefficients between the model predictions and actual observer data are low: 0.22, 0.024, −0.001, and −0.007 for average percentages

Figure 6. Correlation between reaction times and discrimination performance averaged over all observers. Chance performance is at 33.3% correct.
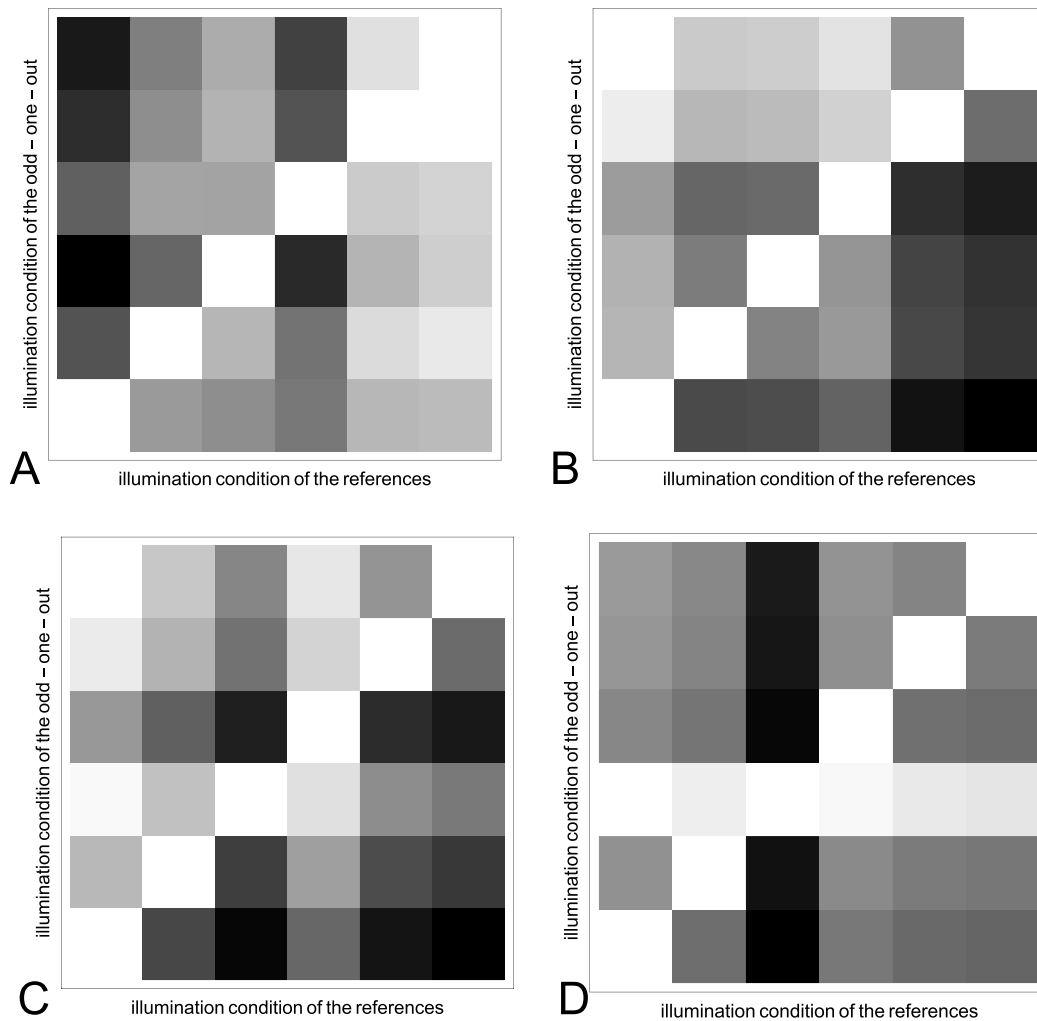
Figure 7. Predictions of performance with the (A) PixelDifferencesModel, (B) MeanLuminanceModel, (C) BrightestLuminanceModel, and (D) LuminanceSkewnessModel. Illumination conditions are ordered in the same way as for Figures 4 and 5. Black means large difference; white means small difference (in arbitrary units).

correct and −0.35, 0.023, 0.070, and 0.097 for average reaction time. A series of *t* tests after Fisher z transformation of the correlation data reveals that there is no statistical difference between the correlation coefficients for these models (*p* values are well above the Bonferroni-corrected threshold of 0.008 for all comparisons). The low correlation coefficients suggests that participants are probably using more specific information to do the task. This is visualized in Figure 8, where we plotted the average percentage correct and the average reaction time for a particular illumination combination against the scaled average pixel differences, mean luminance, brightest luminance, and luminance skewness for that same illumination combination.

Our second model, termed the IlluminationModel, is a bit more sophisticated. It assumes that observers can extract all relevant information about illumination from the scene. The way we model this is to calculate

the difference in pixel values of the images *when all three objects are the same*. Despite this never actually occurring in the experiment, this approach can reveal how much the images of a single object vary between illumination conditions and so how distinguishable the illumination is when it falls on the same object. We calculate these total pixel differences for all three objects (orange, teapot, and tennis ball) separately and average the resulting three values. If we again plot the highest pixel difference in black and the lowest pixel difference in white, we obtain the pattern of results that is shown in Figure 9A. This correlates nicely with the results of our observers (Pearson correlation coefficient of 0.85 for the percentages correct and of −0.85 for the reaction times). A series of *t* tests after Fisher z transformation of the correlation data reveals that these correlation coefficients are indeed higher than the correlations shown in Figure 8 (all *p* values are well below the Bonferroni-corrected threshold of 0.0125).
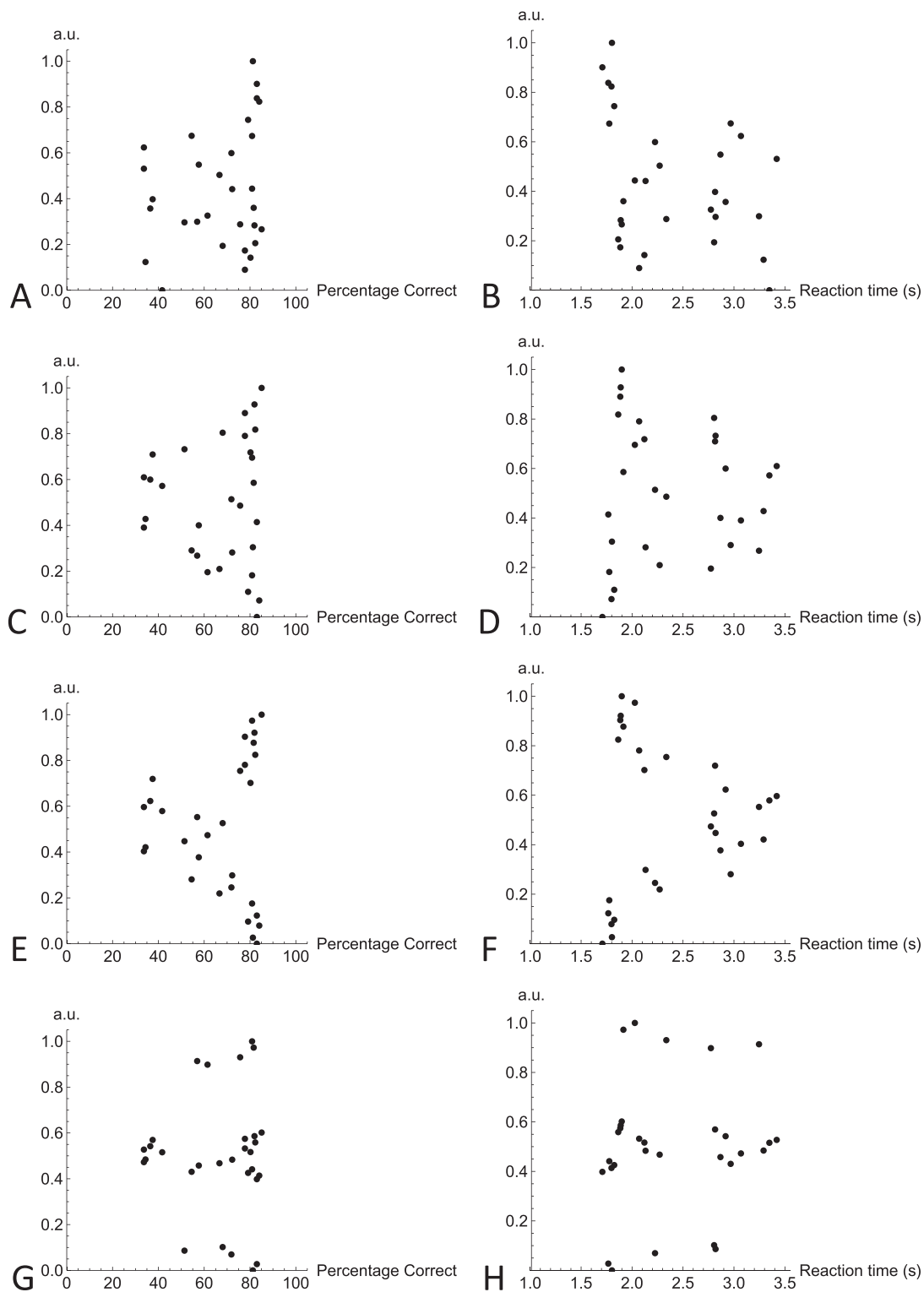
Figure 8. (A) Correlation between discrimination performance averaged over all observers and several image statistics models. (B) Reaction times averaged over all observers and the PixelDifferenceModel. (C) Correlation between discrimination performance averaged over all observers and the MeanLuminanceModel. (D) Reaction times averaged over all observers and the MeanLuminanceModel. (E) Correlation between discrimination performance averaged over all observers and the BrightestLuminanceModel. (F) Reaction times averaged over all observers and the BrightestLuminanceModel. (G) Correlation between discrimination performance averaged over all observers and the LuminanceSkewnessModel. (H) Reaction times averaged over all observers and the LuminanceSkewnessModel.
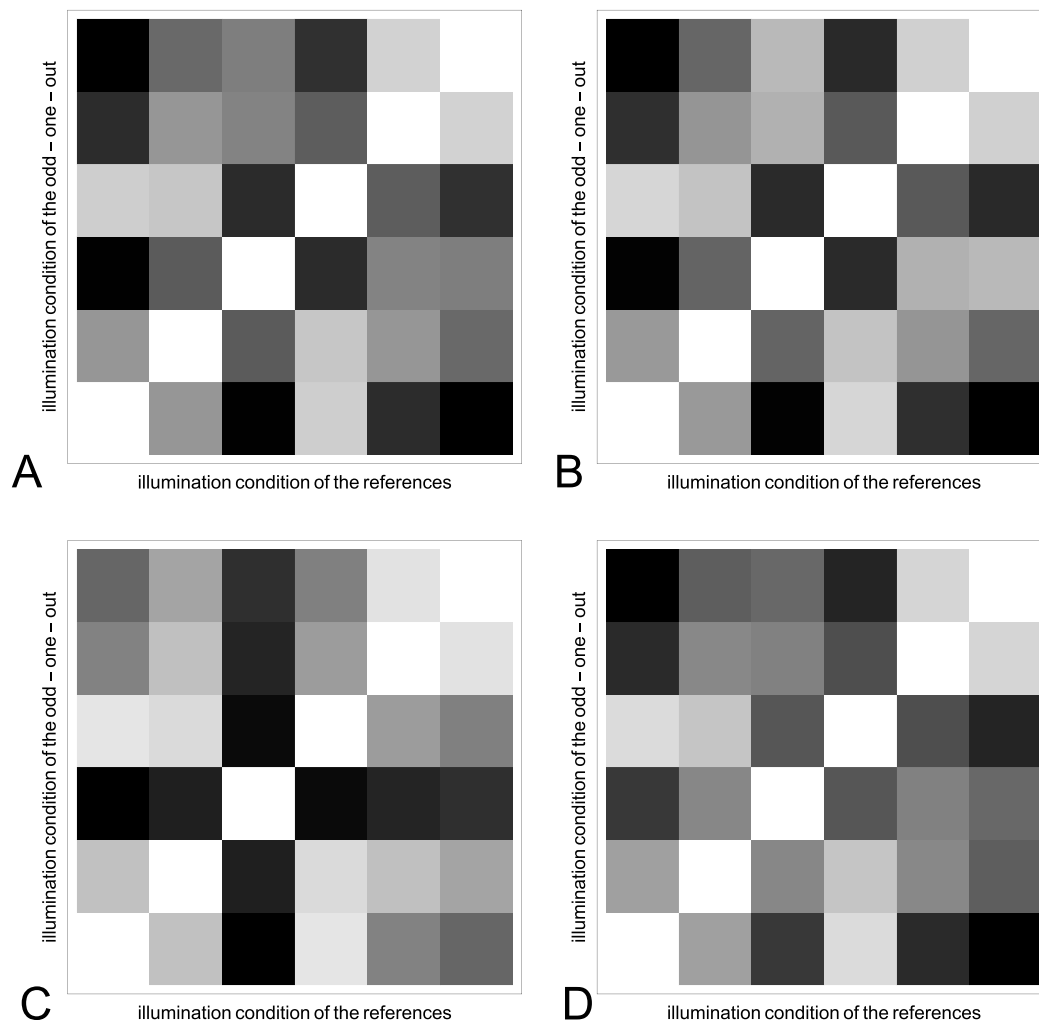
Figure 9. Predictions of performance with (A) IlluminationModel, (B) HighlightModel, (C) ShadingModel, and (D) ShadowModel. Illumination conditions are ordered in the same way as in Figure 3.

The remaining three models use the same method but are restricted to a specific AOI of the image, containing either the highlight on the object (HighlightModel), the shading on the object (ShadingModel), or the shadow cast by the object (ShadowModel). The pattern of results we obtain for these three models is shown in Figures 9B–D.

Clearly, the pattern of results from the shading model correlates less well with the pattern of results we obtained from observers (Pearson correlation coefficient of 0.54 for the percentages correct and of −0.41 for the reaction times). The pattern of results from both the shadow and highlight models correlates highly with the pattern of results we obtained from observers, with the highlight and shadow models correlating best with both the reaction time data (Pearson correlation coefficient of 0.84 for the percentages correct and of −0.90 for the reaction times for highlights, correlation coefficient of 0.87 for the percentages correct and of −0.85 for the reaction times

for shadows). This is visualized in Figures 10 and 11, where we plotted the average percentage correct and the average reaction time for a particular illumination combination against the scaled average image information for that same illumination combination. A series of t tests after Fisher z transformation of the correlation data reveals that there is no statistical difference between the correlation coefficients for the HighlightModel and the ShadowModel ($p = 0.75$ for percentage correct, $p = 0.50$ for reaction times). However, the correlation coefficient for the Shading-Model is significantly lower than both ($p = 0.025$ and $p = 0.011$ for percentage correct and $p = 0.0002$ and $p = 0.0020$, respectively, for reaction times).

Eye Fixations

Comparing our different models with the observer data suggests that observers used mainly highlights
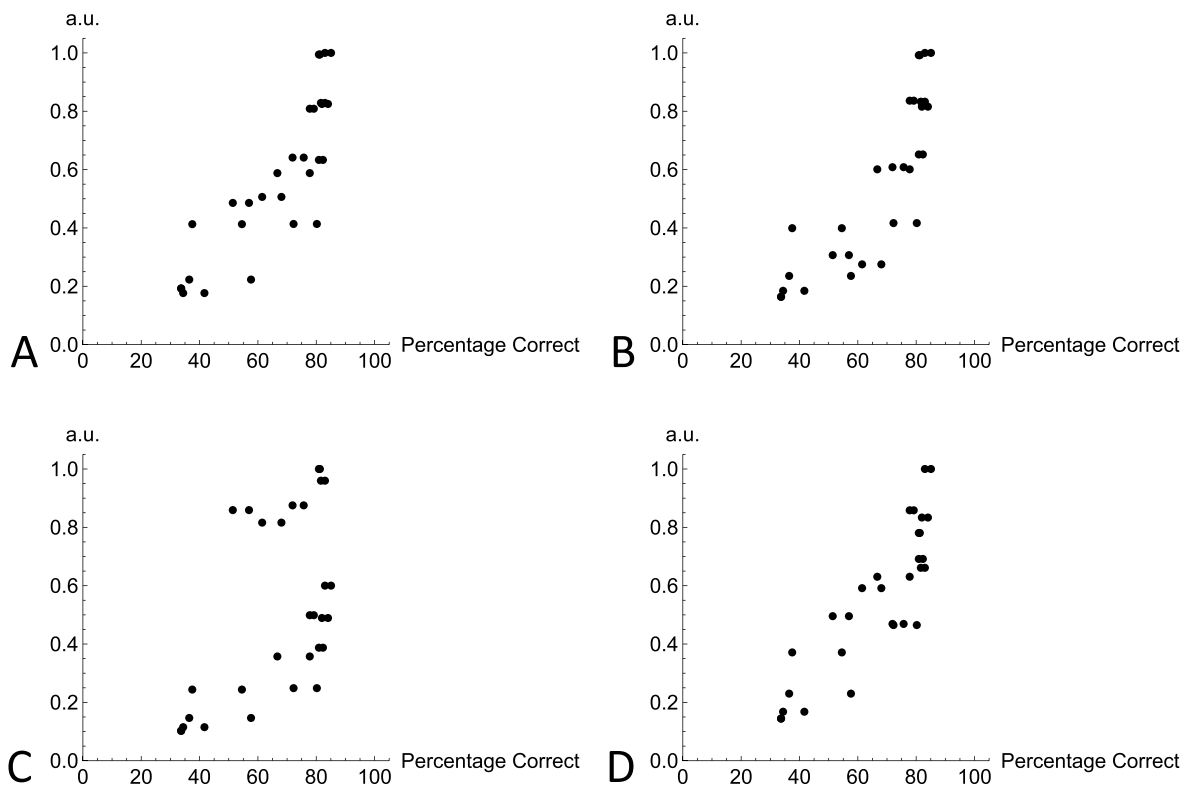
Figure 10. Correlation between discrimination performance averaged over all observers and the IlluminationDifferencesModel for (A) All AOIs, (B) highlight AOI, (C) shading AOI, and (D) shadow AOI.
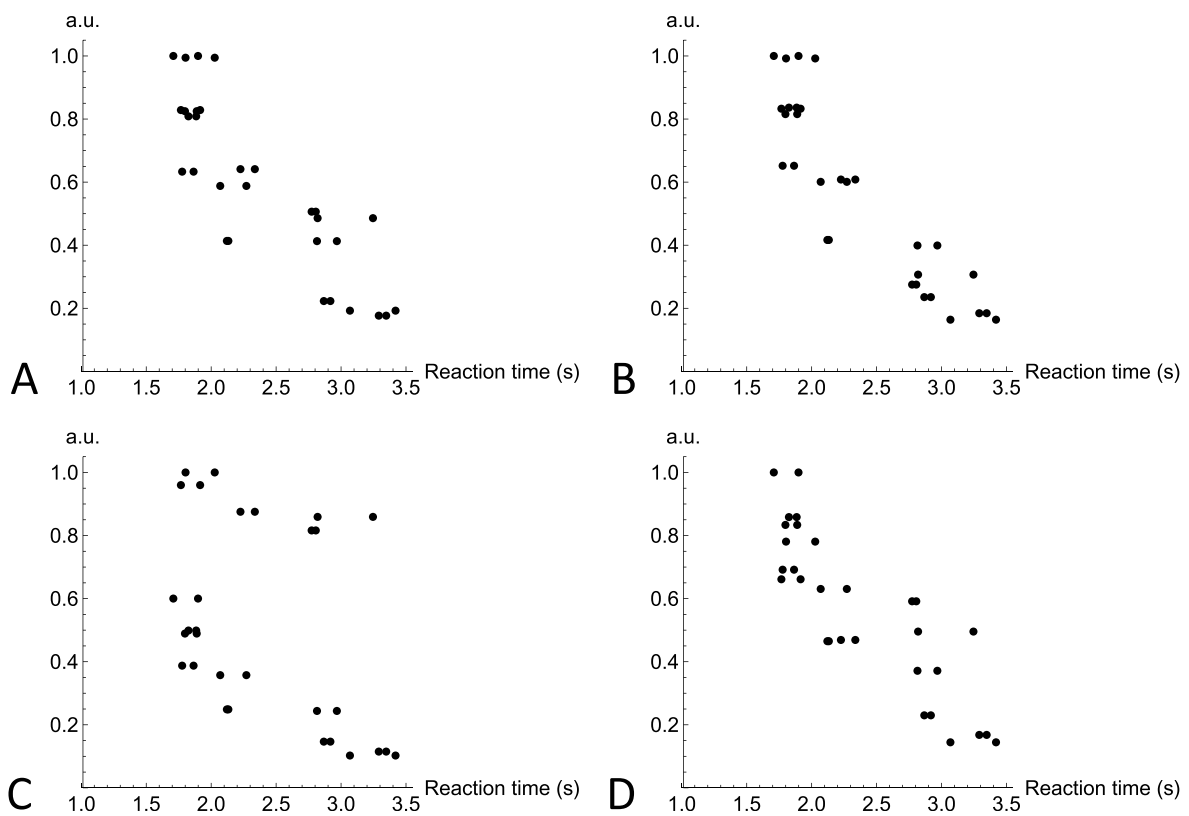

Figure 11. Correlation between reaction times averaged over all observers and the IlluminationDifferencesModel for (A) all AOIs, (B) highlight AOI, (C) shading AOI, and (D) shadow AOI.

and cast shadows to distinguish the illumination conditions. To see whether we can observe such a preference for highlights and shadows in looking behavior, we analyzed observers' eye fixations. We divided the stimuli roughly into three AOIs: the top part of the images, containing the highlights; the middle part, containing most of the shading; and the lowest part, containing most of the shadows. About 82% of fixations were inside the image regions. We discarded all fixations outside of the image regions. We found that of all fixations within the images, 64% were in the shadow AOI, 27% in the shading AOI, and 9% in the highlight AOI. Analyzing the different objects separately revealed that this distribution did not depend on which object the observers were looking at.

Although this gives us some quantitative measure that suggests observers fixated the shadows more often, the more interesting information is contained in the location and the order of the first five fixations. We plotted heat maps of first through fifth fixations for all observers in Figure 12. From this, we can clearly see that, apart from the first fixation that is located near the fixation cross, participants look mainly at the cast shadows of the three objects in a systematic way, as if they are comparing them (as well as near the fixation cross). Apparently, they hardly fixate in the locations containing shading and highlights.

Discussion

We investigated how well observers are able to distinguish the diffuseness of the light and the number of light sources in real photographed scenes and what kind of stimulus information is most important for such a task by asking observers to distinguish different illumination conditions in an odd-one-out task.

Our results show that participants performed above chance for most conditions, and there are systematic variations in performance over conditions. The differences in performance are not predicted well by several simple image statistics such as the average pixel differences, the mean luminance differences, the differences between the brightest luminance in the scenes, and differences in the skewness of the luminance histograms. However, the variations in performance are predicted well by a model that assumes that observers can extract information about illumination from the scene, using the differences in image structure in same-object comparisons. The advantage of such a model is that we can also look at parts of the images containing different types of information about the illumination, such as the highlights, shading, and cast shadows. The resulting

simulations show that participants' performance can be explained by assuming they rely either on the cast shadows or highlights or both to compare illumination conditions. This is in line with the fact that in our rather simple stimuli, the correspondence of the cast shadow to the object is obvious, making the shadow a reliable source of information (Mamassian, 2004), apparently not only for complex scenes and light source direction and intensity (Casati, 2004; Koenderink et al., 2003; Koenderink et al., 2004; Xia et al., 2014) but also for matching illuminant diffuseness and source separation.

When we look at the reaction times, we see that the pattern of results is similar to that of performance: Higher performance correlates with faster reactions. When there is a larger difference in highlight structure or cast shadows of the images, the reaction times are shorter. Shading information is less predictive of participants' reaction times and accuracy. This might be due to the fact that although the images are of real objects, the context they provide is rather sparse and there are very few shape cues (O'Shea et al., 2010; Ostrovsky et al., 2005). Boyaci et al. (2006) showed that all three types of information are used in their scenes; they could rule out the use of shading information in only one participant. However, we show here that it is rather unlikely that using shading information has any advantage for our participants in both reaction time and accuracy. Whether it is likely that participants use cast shadow and highlight information equally is harder to say, because in our models, there is a high correlation (0.90) between the information contained in highlights and shadows. The information contained in the shading does not correlate as well with the highlights and the shadows (0.53 and 0.55, respectively).

There could be a geometrical explanation for the perceptual difference we find between the use of information between highlights, shading, and cast shadows. From cast shadows and highlights, we can infer mathematically higher-order information about the light field than from shading; shading is due to the diffuse part of the reflectance only. Ramamoorthi and Hanrahan (2001) showed that the diffuse part of the reflectance can be described by the second-order spherical harmonics description of the light field. Thus, from the shading, we can infer only the 0th-, first-, and second-order spherical harmonics components of the light field, which are related to intensity, average direction and strength of the light vector, diffuseness (Xia, Pont, & Heynderickx, 2016a, 2016b), and the squash tensor (see Mury, Pont, & Koenderink 2009). Cast shadows and highlights do allow the inference of higher-order contributions (Bunteong & Chotikakamthorn, 2016; Sato, Sato, & Ikeuchi, 2003). Because we varied both diffuseness of the light source

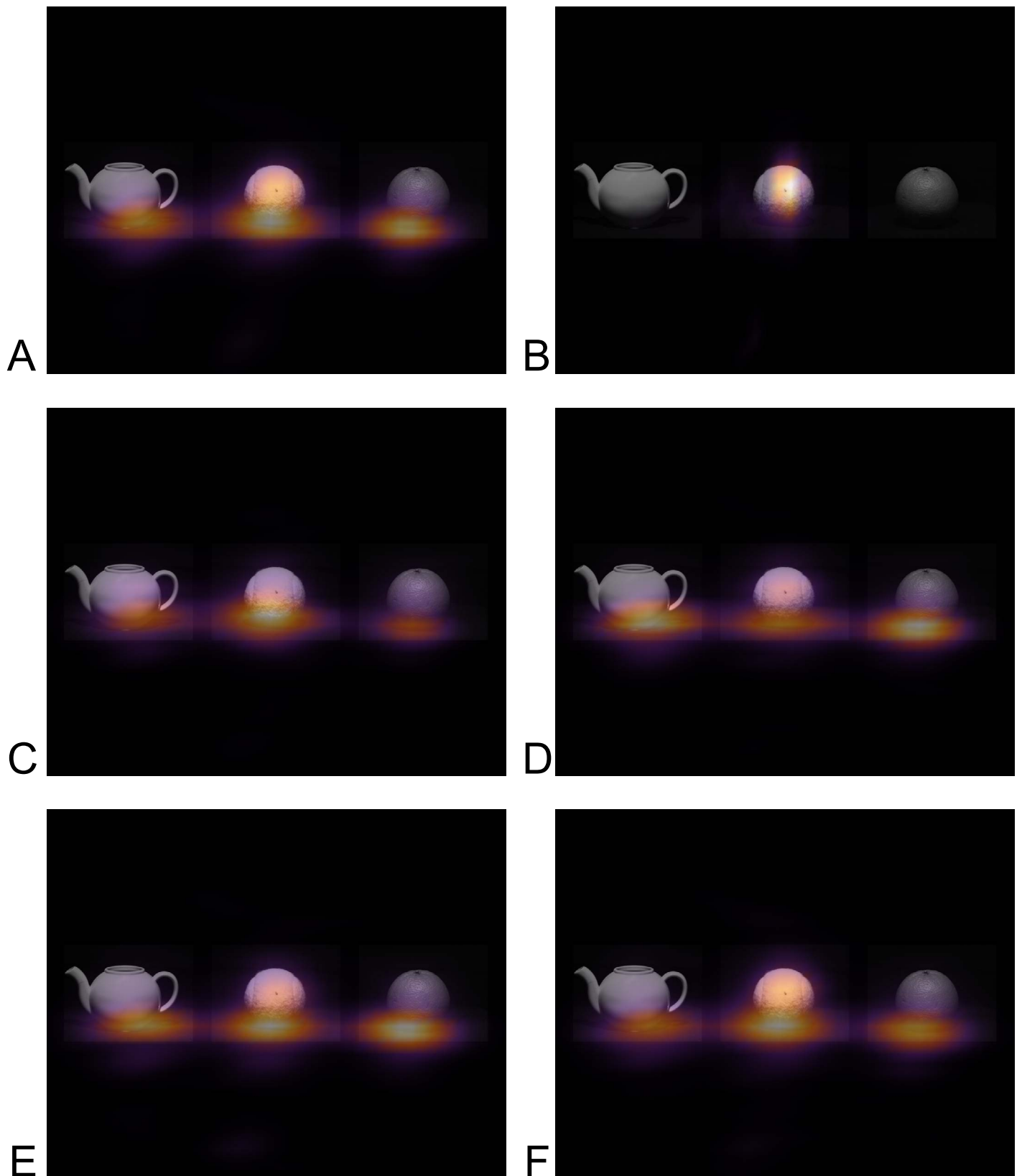Figure 12. Eye fixation density plots combined for all participants. (A) All fixations. (B) All first fixations. (C) All second fixations. (D) All third fixations. (E) All fourth fixations. (E) All fifth fixations. Note that the order of the objects in this example is arbitrary; we combined all eye fixations of the 18 different layouts per illumination combination. Analyzing them separately yields similar results.

and the number of light sources, and the latter distinction requires higher-order information, this might explain our perceptual results.

Interestingly, the eye-fixation data reveal that participants primarily look at the cast shadows in favor of shading and highlights. First through fifth fixations reveal that after the first fixation near the fixation point, observers mainly move their eyes between the shadows of the three objects, seemingly comparing them. This supports our notion that shading information is hardly used in this case, but it seems contradictory to the fact that highlight information in the image also predicts both the performance and reaction time data well and also contradicts data from Boyaci et al. (2006). However, it could be the case that observers did not need to make many fixations to extract the information from the highlights but were able to extract this high-contrast information directly at fixation. This is in accordance with results from Sharan et al. (2009, 2014) that show (among other things) that gloss can be categorized within 40 ms.

Conclusions

We show that observers' performance on illumination matching in simple real scenes is best predicted by a model that uses only highlights and cast shadow information about the illuminant. This was confirmed by fixation data; observers fixated mostly in the cast shadow areas. However, the number of eye fixations in highlight areas is rather low. This could be due to the fact that highlight information might be readily available to the observers, even at larger eccentricities. We conclude that observers rely mainly on cast shadows and possibly also on highlights to distinguish illumination conditions that vary in low-level aspects such as number of sources and diffuseness. This makes sense from a geometrical point of view: Higher-order information on the light field is available only from shadows and highlights, not from shading.

Keywords: illumination perception, light perception, material perception, highlights, shadows, shading, eye fixations, reaction times

Acknowledgments

Commercial relationships: none.
Corresponding author: Susan F. te Pas.
Email: s.tepas@uu.nl.
Address: Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, the Netherlands.

References

Adelson, E. H., & Pentland A. P. (1996). The perception of shading and reflectance. In D. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 409–423). New York: Cambridge University Press.

Anderson, B. L. (2011). Visual perception of materials and surfaces. *Current Biology*, *21*, R978–R983.

Belhumeur, P. N., Kriegman, D. L., & Yuille, A. L. (1999). The bas-relief ambiguity. *International Journal of Computer Vision*, *35*, 33–44.

Blake, A., & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature*, *343*, 165–168.

Boyaci, H., Doerschner, K., & Maloney, L. T. (2006). Cues to an equivalent lighting model. *Journal of Vision*, *6*(2):2, 106–118, doi:10.1167/6.2.2. [PubMed] [Article]

Boyaci, H., Fang, F., Murray, S. O., & Kersten, D. (2007). Responses to lightness variations in early human visual cortex. *Current Biology*, *17*, 989–993.

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision*, *3*(8):2, 541–553, doi:10.1167/3.8.2. [PubMed] [Article]

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Braje, W. L., Legge, G. E., & Kersten, D. (2000). Invariant recognition of natural objects in the presence of shadows. *Perception*, *29*, 383–398.

Bunteong, A., & Chotikakamthorn, N. (2016). Light source estimation using feature points from specular highlights and cast shadows. *International Journal of Physical Sciences*, *11*, 168–177.

Casati, R. (2004). The shadow knows: A primer on the informational structure of cast shadows. *Perception*, *33*, 1385–1396.

Doerschner, K., Boyaci, H., & Maloney, L. T. (2007). Testing limits on matte surface color perception in three-dimensional scenes with complex light fields. *Vision Research*, *47*, 3409–3423.

Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004).

Statistical characterization of real-world illumination. *Journal of Vision*, *4*(9):11, 821–837, doi:10.1167/4.9.11. [PubMed] [Article]

Ferwerda, J. A., Pellacini, F., & Greenberg, D. P. (2001). Psychophysically based model of surface gloss perception. *Proceedings of SPIE Human Vision and Electronic Imaging VI*, *4299*, 291, doi:10.1117/12.429501.

Fleming, R. W. (2012). Human perception: Visual heuristics in the perception of glossiness. *Current Biology*, *22*, R865–R866

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research*, *94*, 62–75.

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real world illumination and the perception of surface reflectance properties. *Journal of Vision*, *3*(5):3, 347–368, doi:10.1167/3.5.3. [PubMed] [Article]

Geusebroek, J. M., Burghouts, G. J., & Smeulders, A. W. M. (2005). The Amsterdam library of object images. *International Journal of Computer Vision*, *61*, 103–112.

Gilchrist, A. L. (1994). *Lightness, brightness, and transparency*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hooge, I. T. C., & Camps, G. (2013). Scan path entropy and arrow plots: capturing scanning behavior of multiple observers. *Frontiers in Psychology*, *4*, 1–10, dx.doi.org/10.3389/fpsyg.2013.00996.

Hurlbert, A. (1999). Colour vision: Is colour constancy real? *Current Biology*, *9*, R558–R561.

Kartashova, T., de Ridder, H., te Pas, S. F., Schoemaker, M., & Pont, S. C. (2015). The visual light field in paintings of Museum Prinsenhof: Comparing settings in empty space and on objects. *SPIE Electronic Imaging*, 93941M–93941M-10

Kartashova, T., Sekulovski, D., de Ridder, H., te Pas, S. F., & Pont, S. C. (2016). Global structure of the visual light field and its relation to the physical light field. *Journal of Vision*, *16*(10):9, 1–16, doi:10.1167/16.10.9. [PubMed] [Article]

Koenderink, J. J., Pont, S. C., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2007). The visual light field. *Perception*, *36*, 1595–1610.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., te Pas, S. F., & Pont, S. C. (2003). Illumination direction from texture shading. *Journal of the Optical Society of America A*, *20*, 987–995.

Koenderink, J. J., van Doorn, A. J., & Pont, S. C. (2004). Light direction from shad(ow)ed random Gaussian surfaces. *Perception*, *33*, 1405–1420.

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences, USA*, *96*, 307–312.

Ling, Y., & Hurlbert, A. (2004). Color and size interactions in a real 3D object similarity task. *Journal of Vision*, *4*(9):5, 721–734, doi:10.1167/4.9.5. [PubMed] [Article]

Maloney, L., Gerhard, H., Boyaci, H., & Doerschner, K. (2010). Surface color perception and light field estimation in 3D scenes. In L. R., Harris, & M. Jenkins (Eds.), *Vision in 3D environments* (pp. 65–88). Cambridge, UK: Cambridge University Press.

Mamassian P. (2004). Impossible shadows and the shadow correspondence problem. *Perception*, *33*, 1279–1290.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*, B1–B9.

Marlow, P. J., Kim, J., & Anderson, B. L. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, *22*, 1909–1913.

Morgenstern, Y., Geisler, W. S., & Murray, R. F. (2014). Human vision is attuned to the diffuseness of natural light. *Journal of Vision*, *14*(9):15, 1–18, doi:10.1167/14.9.15.

Morgenstern, Y., Murray, R. F., & Harris, L.R. (2011). The human visual system's assumption that light comes from above is weak. *Proceedings of the National Academy of Sciences, USA*, *108*, 12551–12553.

Motoyoshi, I., Nishida, S. Y., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, *447*, 206–209.

Mury, A. A., Pont, S. C., & Koenderink, J. J. (2009). Structure of light fields in natural scenes. *Applied Optics*, *48*, 5386–5395.

Nishida, S., & Shinya, M. (1995). Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A*, *15*, 2951–2965.

Obein, G., Knoblauch, K., & Vienot, F. (2004). Difference scaling of gloss: Non-linearity, binocularity, and constancy. *Journal of Vision*, *4*(9):4, 711–720, doi:10.1167/4.9.4.

O'Shea, J. P., Agrawala, M., & Banks, M. S. (2010). The influence of shape cues on the perception of lighting direction. *Journal of Vision*, *10*(12):21, 1–21, doi:10.1167/10.12.21. [PubMed] [Article]

Ostrovsky, Y., Cavanagh, P., & Sinha, P. (2005). Perceiving illumination inconsistencies in scenes. *Perception*, *34*, 1301–1314.

Pelli, D. G. (1997). The Videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision 10*, 437–442.

Pont, S. C., & Koenderink, J. J. (2007). Matching illumination of solid objects. *Perception & Psychophysics*, *69*, 459–468.

Pont, S. C., Koenderink, J. J., van Doorn, A. J., Wijntjes, M. W. A., & te Pas, S. F. (2012). Mixing material modes. *Proceedings of the SPIE: Human Vision and Electronic Imaging, 8291*, 82910D, doi:10.1117/12.916450.

Pont, S. C., & te Pas, S. F. (2006). Material-illumination ambiguities and the perception of solid objects. *Perception*, *35*, 1331–1350.

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., & Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision*, *16*(11):2, 1–18, doi:10.1167/16.11.2. [PubMed] [Article]

Ramamoorthi, R., & Hanrahan, P. (2001). On the relationship between radiance and irradiance: Determining the illumination from images of a convex Lambertian object. *Journal of the Optical Society of America A*, *18*, 2448–2459.

Ripamonti, C., Bloj, M., Hauck, R., Mitha, K., Greenwald, S., Maloney, S. I., & Brainard, D. H. (2004). Measurements of the effect of surface slant on perceived lightness. *Journal of Vision*, *4*(9):7, 747–763, doi:10.1167/4.9.7. [PubMed] [Article]

Robilotto, R., & Zaidi, Q. (2004). Limits of lightness identification for real objects under natural viewing conditions. *Journal of Vision*, *4*(9):9, 779–797, doi:10.1167/4.9.9. [PubMed] [Article]

Sato, I., Sato, Y., & Ikeuchi, K. (2003). Illumination from shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*, 290–300.

Schirillo, J. A. (2013). We infer light in space. *Psychonomic Bulletin and Review*, *20*, 905–915.

Sharan, L., Li, Y., Motoyoshi, I., Nishida, S. Y., & Adelson, E. H. (2008). Image statistics for surface reflectance perception. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *25*, 846–865.

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2009). Material perception: What can you see in a brief glance? *Journal of Vision*, *9*(8): 784, doi:10.1167/9.8.784. [Abstract]

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2014). Accuracy and speed of material categorization in real-world images. *Journal of Vision*, *14*(9):12, 1–24, doi:10.1167/14.9.12. [PubMed] [Article]

te Pas, S. F., & Koenderink, J. J. (2004). Visual discrimination of spectral distributions. *Perception*, *33*, 1483–1497.

te Pas, S. F., & Pont, S. C. (2005). A comparison of material and illumination discrimination performance for real rough, real smooth and computer generated smooth spheres. *Proceedings of the ACM SIGGRAPH Symposium on APGV*, 75–83.

Toscani, M., Valsecchi, M., & Gegenfurtner K. R. (2013). Optimal sampling of visual information for lightness judgments. *Proceedings of the National Academy of Sciences, USA*, *110*, 11163–11168.

Toscani, M., Valsecchi, M., & Gegenfurtner K. R. (2017). Lightness perception for matte and glossy complex shapes. *Vision Research*, *131*, 82–95.

Toscani, M., Zdravković, S., & Gegenfurtner, K. R. (2016). Lightness perception for surfaces moving through different illumination levels. *Journal of Vision*, *16*(15):21, 1–18, doi:10.1167/16.15.21. [PubMed] [Article].

van Assen, J. J. R., Wijntjes, M. W. A., & Pont, S. C. (2016). Highlight shapes and perception of gloss for real and photographed objects, *Journal of Vision*, *16*(6):6, 1–14, doi:10.1167/16.6.6. [PubMed] [Article]

van Doorn, A. J., Koenderink, J. J., Todd, J. T., & Wagemans, J. (2012). Awareness of the light field: The case of deformation. *i-Perception*, *3*, 467.

Wiebel, C. B., Toscani, M., & Gegenfurtner, K. R. (2015). Statistical correlates of perceived gloss in natural images. *Vision Research*, *115*, 175–187.

Wiebel, C., Valsecchi, M., & Gegenfurtner, K. R. (2013). The speed and accuracy of material recognition in natural images. *Attention, Perception, & Psychophysics*, *75*(5), 954–966.

Xia, L., Pont, S. C., & Heynderickx, I. (2014). The visual light field in real scenes. *i-Perception*, *5*, 613–629.

Xia, L., Pont, S. C., & Heynderickx, I. (2016a). Light diffuseness metrics I: Theory. *Lighting Research and Technology*, *1477153516631391*, doi:10.1177/1477153516631391.

Xia, L., Pont, S. C., & Heynderickx, I. (2016b). Light diffuseness metrics II: Describing, measuring and visualizing the light flow and diffuseness in 3D spaces. *Lighting Research and Technology*, *1477153516631392*, doi:10.1177/1477153516631392.

Zhang, F., de Ridder, H., & Pont, S. C. (2015). The influence of lighting on visual perception of material qualities. *IS&T/SPIE Electronic Imaging*, 93940Q–93940Q-10.