Vision Research 131 (2017) 82–95

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Lightness perception for matte and glossy complex shapes

Matteo Toscani *, Matteo Valsecchi, Karl R. Gegenfurtner

Abteilung Allgemeine Psychologie, Justus-Liebig-Universität Giessen, Otto-Behaghel-Str. 10F, 35394 Giessen, Germany

ARTICLE INFO

Article history:
Received 21 July 2016
Received in revised form 12 December 2016
Accepted 13 December 2016
Available online 13 January 2017

Keywords:
Material perception
Lightness perception
Image statistics
Shading
Gloss
Three-dimensional shape

ABSTRACT

Humans are able to estimate the reflective properties of the surface (albedo) of an object despite the large variability in the reflected light due to shading, illumination and specular reflection. Here we first used a physically based rendering simulation to study how different statistics (i.e. percentiles) based on the luminance distributions of matte and glossy objects predict the overall surface albedo. We found that the brightest parts of matte surfaces are good predictors of the surface albedo. As expected, the brightest parts led to poor performance in glossy surfaces. We then asked human observers to sort four (2 matte and 2 glossy) objects in a virtual scene in terms of their albedo. The brightest parts of matte surfaces highly correlated with human judgments, whereas in glossy surfaces, the highest correlation was achieved by percentiles within the darker half of the objects' luminance distributions. Furthermore, glossy surfaces tend to appear darker than matte ones, and observers are less precise in judging their lightness. We then manipulated different bands of the virtual objects' luminance distributions separately for glossy and matte surfaces. Modulating the brightest parts of the luminance distributions of the glossy surfaces had a limited impact on lightness perception, whereas it clearly influenced the perceived lightness of the matte objects. Our results demonstrate that human observers effectively ignore specular reflections while evaluating the lightness of glossy objects, which results in a bias to perceive glossy objects as darker.

© 2016 Published by Elsevier Ltd.

1. Introduction

The light reaching the eye from a surface depends on the albedo of the surface, the illumination, the geometry of the surface, and the transmitting medium between the reflecting surface and the eye. In nature, geometries or illuminants typically cause luminance variations in the light reflected from the surface (i.e. shading), even when the surface is made of a single material. Apparently, human observers are able to perceive both lightness, defined as the apparent reflectance of an object's surface (not affected by shading), and brightness, defined as the apparent luminance, at the same time (Arend & Spehar, 1993). We can perceive brightness because otherwise we would not be able to perceive shading at all. Lightness constancy is the ability of our visual system to recover the albedo (diffuse reflectivity) of an object's surface despite changes in the environmental conditions. This task is far from trivial because surfaces with different albedos (e.g. one dark and one light surface) can produce luminance distributions that overlap to a large extent, due to the interaction between the surface geometry and the illuminant (shading). It is therefore interesting to study how the visual system discounts shading. In these terms, one potential contribution to lightness constancy is the ability to tell shading and albedo apart.

Several investigators proposed that in order to recover surface albedo, the visual system explicitly estimates and discounts the contributions of illumination and geometry to the observed luminances (e.g. Marr, 1982; Pizlo, 2001; Poggio & Koch, 1985; Poggio, Torre, & Koch, 1985). This approach is referred to as *inverse optics*. An alternative theoretical approach, proposes that the visual system uses simple image statistics to bypass this problem and estimate surface albedo directly (see for review: Fleming, 2014; Thompson, Fleming, Creem-Regehr, & Stefanucci, 2011). This *image statistics* approach is motivated by the sheer impossible difficulty of estimating the individual factors when naturalistically complex geometries are concerned. In fact, the majority of studies about lightness perception are based on simplified stimuli: flat matte surfaces placed on a single plane under diffuse illumination (for an overview, see Maloney & Brainard, 2010). Under these simplified conditions, edges were proposed as the crucial information to estimate the relative albedo of coplanar surfaces (Cornsweet, 1970; Land & McCann, 1971). Nishida and Shinya (1998) showed

* Corresponding author at: Justus-Liebig-Universität Gieβen, Fachbereich 06, Psychologie und Sportwissenschaft, Abteilung Allgemeine Psychologie, Otto-Behaghel-Straβe 10F, 35394 Gieβen, Germany

E-mail address: matteo.toscani@psychol.uni-giessen.de (M. Toscani).

http://dx.doi.org/10.1016/j.visres.2016.12.004
0042-6989/© 2016 Published by Elsevier Ltd.

that more complex geometries can lead observers to produce large errors in matching the diffuse and specular reflectance components between different shapes (Nishida & Shinya, 1998). The pattern of errors in their study suggested that the reflectance matches were based on the similarity of the luminance histograms between the images. However, additional complexity can also provide additional cues to lightness perception. For instance, in simple conditions, when a matte planar surface is viewed in isolation, albedo and illumination are impossible to distinguish and isolated flat surfaces are perceived as white (Gelb, 1929). Nonetheless, observers can – to some extent – judge the lightness of relatively more complex surfaces, such as stucco, even if they are presented in isolation (Sharan, Li, Motoyoshi, Nishida, & Adelson, 2008). In these experiments, Sharan and colleagues tested lightness perception using photographs of real planar surfaces of matte and glossy materials, uniform in albedo. The additional complexity given by the mesostructure and the specular reflection enabled the observers to judge the albedo, despite the fact that the surfaces were presented in isolation. The authors observed that it is possible to predict lightness judgments based on the luminance histogram. Namely, they found that measured albedo and lightness (perceived albedo) correlate negatively with skewness, standard deviation and the 90th percentile of the surfaces' luminance distribution. In a companion paper, Motoyoshi and colleagues (Motoyoshi, Nishida, Sharan, & Adelson, 2007) showed that skewness correlates negatively with albedo and perceived lightness, and positively with the presence of specular reflections and perceived gloss.

We previously found that the highest percentiles of the luminance distributions of matte objects are particularly diagnostic for their albedo, and human observers tend to base their lightness judgments on them (Toscani, Valsecchi, & Gegenfurtner, 2013; Toscani, Valsecchi, & Gegenfurtner, 2015). It is not yet clear whether the visual system applies the same strategy when exposed to glossy surfaces. Intuitively, the brightest parts of the luminance distributions of glossy surfaces should be contaminated to a large extent by specular reflection. Consequently, specular highlights could be seen as a source of noise on the diffuse reflection, and a different heuristic could be used. In the present study we investigate what aspects of the luminance distributions are related to surface albedo and to human lightness judgments (separately for glossy and matte surfaces). More generally, we want to study how the presence of specular reflection impacts lightness perception, in terms of precision and appearance.

If lightness perception is indeed based on the brightest parts of the luminance distributions of both matte and glossy surfaces, the latter ones should appear lighter. If the specular highlights are discounted and the lightness judgments are based on the remaining parts of the distribution, glossy surfaces should appear darker, unless some active compensation takes place. However, comparing the lightness of glossy and matte surfaces is not a trivial task. Even equating the diffuse reflectance across gloss levels is not a trivial problem from a physical point of view. Diffuse reflectance can be defined as (1) the diffuse flux in proportion to the incident illumination, or (2) in proportion to a component of incident illumination that discounts the illumination lost through specular reflection. This distinction is crucial because one difference between diffuse and specular reflection is that in the former case, the light is reflected in all directions, whereas in the latter the direction depends on the surface normal and on the illuminant direction (Hero's law, see Heath, 1921). For this reason, specular highlights appear only reflected from the points of a glossy surface that project toward the point of view of the observer, and when the observer moves, the highlights appear on a different part of the surface. This implies that the light is specularly reflected by the whole surface, but the specular highlights of a certain region of the surface are visible only from the appropriate point of view. If we commit to the second definition of diffuse reflectance, given that part of the incoming light is specularly reflected from every point of the surface, glossy surfaces present areas where the radiance reflected is actually lower (lowlights, see Kim, Marlow, & Anderson, 2012), as compared to a matte surface with the same diffuse reflection component, and brighter areas where specular and diffuse reflection add when reaching the retina (highlights).

With respect to the precision of lightness judgments, we expect observers to be worse in judging glossy surfaces. This is because specular highlights tend to appear in the proximity of the luminance maxima in diffuse shading (Fleming, Torralba, & Adelson, 2004; Koenderink & van Doorn, 1980). Since those luminance maxima are the most informative about surface albedo, having them contaminated by specular highlights should reduce the precision of lightness judgments.

Here, in our first experiment we used a physically based rendering software (radiance – developed by Ward (1994)) to simulate a set of tridimensional models of objects, both with glossy or matte reflectance, and under different naturalistic illuminants. We then used a classification approach to assess to what extent each percentile of the surface luminance distribution predicts the surface albedo (Wiebel, Toscani, & Gegenfurtner, 2015). Similar to our previous study about matte surfaces (Toscani et al., 2013, 2015), we focused our analysis on the percentile statistics because they are directly available to the observer as luminance of a given section of the object surface, whereas for instance the mean luminance might not be represented in the image at all in the case of object images with very bimodal luminance histograms. We repeated this analysis on a smaller set of rendered scenes where we used a reduced set of "blobby" shapes (described later in detail). With this reduced set of shapes, we tested human participants in a lightness ranking task, aiming to compare their performance with the simulation results. We found the ranking task to be more natural than a standard lightness matching, and thus preferable given that lightness and color judgments are particularly sensitive to task instructions (Arend & Spehar, 1993; Schneider & von Campenhausen, 1998). We used the ranking results to study the importance of the different percentiles of the surfaces luminance distributions on lightness perception, and related this result with the one from the reflection simulations, similar to what we previously did with matte surfaces (Toscani et al., 2013, 2015). In a last experiment, we manipulated different bands of the surface luminance histograms to study the causal impact of the different percentiles on lightness perception (separately for gloss and matte surfaces).

2. Simulation of natural objects

We aimed to find out which aspects of the luminance distributions of complex surfaces are good predictors for surface albedo, separately for gloss and matte surfaces. The brightest parts of the luminance distributions of glossy surfaces are likely to be contaminated by specular reflections, which would constitute a source of noise in the estimation of the diffuse reflection. Conversely, the brightest parts of matte surfaces are the most informative about surface albedo (Toscani et al., 2013, 2015). Here we used a classification algorithm to study how the different percentiles of the surface luminance distributions perform in predicting the surface albedo. For the sake of generality, in our simulations we used a large collection of different tridimensional shapes rendered with several orientations, from several viewpoints, and embedded in several different light fields.

2.1. Methods

Renderings: We created our simulated scenes in an analogous way as Wiebel et al. (2015). We rendered 83 different virtual objects each under 600 random conditions, using the software RADIANCE (Ward, 1994) through the MATLAB-based RenderToolbox3 (Heasly, Cottaris, Lichtman, Xiao, & Brainard, 2014). Only one object was placed in the centre of every rendered scene. The orientation of the object in the scene was randomized by randomly rotating it on its axes. The point from which the scene was viewed was randomly sampled from a sphere surrounding the object. We rendered a total of 51,600 scenes, split into three groups of 17,200. The first group of scenes contained matte objects rendered with a Lambertian reflectance model. The second and the third groups included glossy objects rendered with a Ward model (Ward, 1992), with increasing importance of the specular component. The diffuse reflectance components were randomly chosen for every virtual surface by sampling them from a beta distribution, based on the results of Attewell and Baddeley (2007). These authors showed that the distribution of natural surface reflectances is best approximated by a beta distribution with parameters a = 1.29 and b = 2.30. The specular surfaces were defined by the two classic parameters of the Ward model; one representing the magnitude of the specular component (i.e. specularity; set to 0.2 for the second group and 0.25 for the third group) and the other the "spread" or blur of the specular reflection (i.e. roughness; set to 0.1). To simulate naturalistic illuminants, we used the 9 Debevec's light probes (Debevec, 2008). For every scene, one of these light probes was randomly chosen as the virtual light field (Fig. 1). The three-dimensional models were chosen from the following internet free repositories: the Turbosquid (http://www.turbosquid.com) license free repository, the Stanford Computer Graphics Laboratory (http://graphics.stanford.edu/data/3Dscanrep), and the AIM@SHAPE shape repository (http://visionair.ge.imati.cnr.it/ontologies/shapes). We selected models that exhibited a large range of diversity and complexity, from simple geometrical shapes, such as a cube, to complex three-dimensional figures, like animals.

Classification analysis: We rendered a perfectly dark version of each surface in each position to generate a mask for segmenting it from its background when calculating the luminance histograms. From each luminance distribution, we extracted the percentile statistics and used them to classify the images according to their albedo (diffuse reflection coefficient used in the renderings). We used a linear binary classifier approach, where we divided the images into dark or light albedo by a median split (similarly to Sharan et al., 2008). Classification was done for each individual percentile. We trained the classifier on the whole set of images, leaving out one image at a time. The excluded image was then tested with the leave-one out linear classifier. All classification analyses were done with the classification routines implemented in the "classify" function of the statistics toolbox for MATLAB (version R2015a, http://www.mathworks.com).

2.2. Results

Fig. 2 represents the classification performance, for matte surfaces (red curve) and the two groups of glossy surfaces with lower and higher specularity (green and blue, respectively). Performance increases monotonically as a function of the luminance percentile rank in the matte surface case, similar to our previous results (Toscani et al., 2013, 2015). On the other hand, for both of the glossy groups, performance is relatively stable until approximately the 80th percentile and only a few percent points lower than the

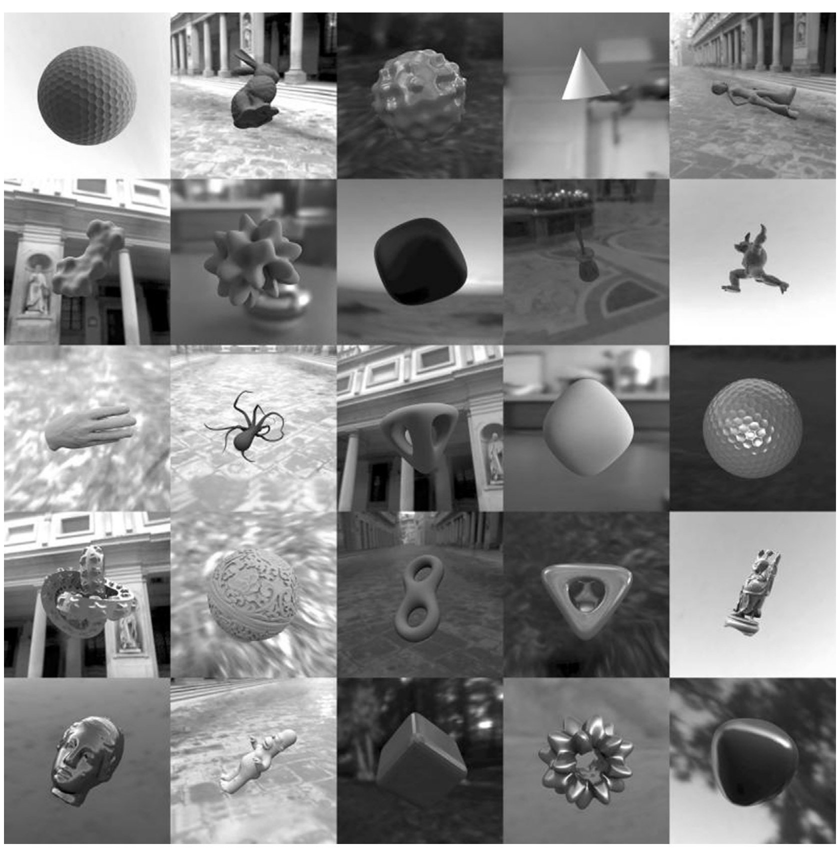

Fig. 1. Examples of the rendered scenes. We selected models that exhibited a large range of complexity, from simple geometrical shapes, such as a cube, to complex three-dimensional figures, like animals. We rendered 83 different virtual objects each under 9 different illumination fields. The figure shows twenty-five random examples of the rendered scenes.

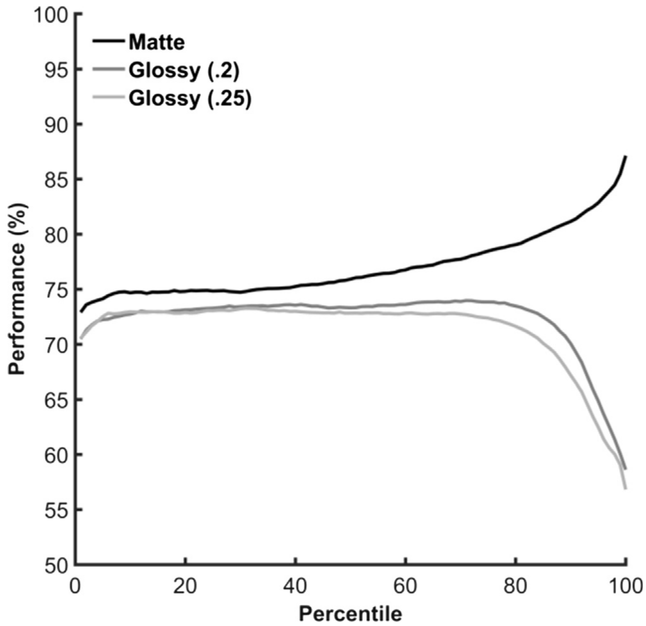

Fig. 2. Albedo classification results for each percentile of matte and glossy surfaces. Performance is represented on the y-axis for each percentile (represented on the x-axis). The solid lines represent the classification performance for matte (black) and glossy (gray, specularity = 0.2; light gray, specularity = 0.25) surfaces. For the matte surfaces, the performance is increasing monotonically as a function of percentile rank, similar to our previous results (Toscani et al., 2013). For both levels of specularity (0.2 and 0.25), for glossy surfaces, performance is relatively stable and comparable to the one obtained with matte surfaces until approximately the 80th percentile, after which it decreases, reaching almost chance levels for the very high percentiles.

performance obtained with matte surfaces, but decreases to almost chance level for the very high percentiles. This result is consistent with the idea that specular reflections are contaminating the brightest parts of a surface (Fleming et al., 2004; Koenderink & van Doorn, 1980), impairing performance.

2.3. Discussion

It appears that while relying on the brightest parts of the surface as an estimate of surface lightness is a viable strategy in the case of matte objects (Toscani et al., 2013, 2015), the same strategy would be suboptimal in the case of glossy objects. In order to investigate whether humans are able to adapt their strategy to the task, in a follow-on experiment we asked a group of participants to rank a set of surfaces in terms of lightness.

3. Simulation with artificial objects

We repeated the classification analysis on a new, more homogeneous set of shapes defined by two relatively simple parameters. This allows for a better understanding and standardization of the results of both simulation and psychophysical testing, also because this family of shapes has been extensively used in vision research before (e.g. Adams, Kerrigan, & Graf, 2016; Cholewiak & Fleming, 2013; Cholewiak, Kunsberg, Zucker, & Fleming, 2014; Cholewiak, Vergne, Kunsberg, Zucker, & Fleming, 2015; Fleming et al., 2004; Muryy, Fleming, & Welchman, 2016; Muryy, Welchman, Blake, & Fleming, 2013; Norman, Todd, & Orban, 2004).

3.1. Methods

Renderings: We rendered our artificial shapes embedded in the artificial scene that we planned to use for behavioral experiments. In a single scene, two surfaces were always matte, and two glossy (Fig. 3). We rendered the four surfaces with four different albedos (40%, 44%, 48% and 52%). These values were chosen after a pilot experiment in order to yield sensibly over chance (but not perfect) performance. We rendered our scenes using the software RADIANCE (Ward, 1994) through the MATLAB-based RenderToolbox3 (Lichtman, Xiao, & Brainard, 2007). The matte surfaces were again rendered according to a Lambertian reflection model, whereas the glossy surfaces were rendered according to the Ward model (Ward, 1992). The specularity parameter was set to 0.2, and the roughness parameter was set to 0.1. We used "The Uffizi Gallery" light probe as the virtual illuminant for all the scenes. We are confident that this choice did not restrict the generality of our results. Indeed human observers can accurately match the diffuse component of a sphere rendered under two different illuminations (Olkkonen & Brainard, 2010) and when matching the diffuse component of two different shapes under two different illuminations the errors were rather small (Olkkonen & Brainard, 2011).

Shapes: A set of "blobby" objects were generated using sinusoidal perturbations of spheres (Cholewiak & Fleming, 2013; Cholewiak et al., 2014, 2015). This sinusoidal perturbation was characterized by its sine frequency and amplitude. We used a set of 16 shapes, defined by four levels of amplitude and four levels of frequency (Fig. 4). The generation algorithm starts with a sphere, and recursively applies 5 sinusoidal perturbations to its vertexes. The perturbation Amplitude ranged between ±[2,4,6,8]% of the sphere radius. The Frequency of each perturbation was defined in terms of the number of cycles of the wave within the sphere (n = 1, 3, 5, 7), and is intuitively understood in terms of the number of "corners" the object has (Muryy et al., 2013).

Scene composition: Four shapes were laying on a marble table in every scene, two glossy and two matte (Fig. 3). Shape Frequency and Amplitude, Specularity and position in the scene were balanced across the full set of rendered scenes. This ensured a certain generality to the simulation results, despite the simplicity of the objects. In each scene, two shapes had one of the two higher amplitude levels and two had one of the two lower amplitude levels. Similarly, two shapes had one of the two higher spatial frequency levels and one of the two lower spatial frequency levels. Frequency and Amplitude are crossed within each scene, thus resulting in all four combinations: 1) one shape with high amplitude and high

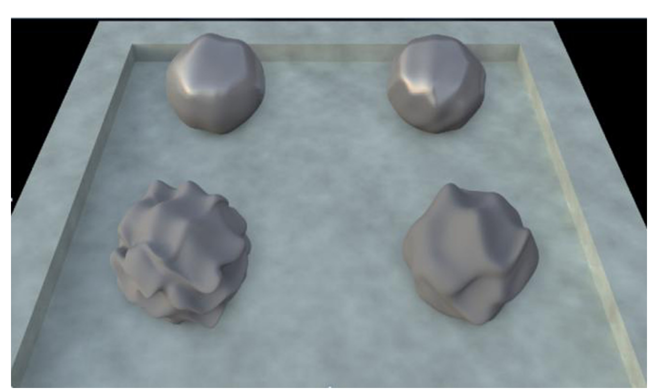

Fig. 3. Scene example. Four shapes were always present in the scene, two glossy and two matte. The table was rendered totally matte with the diffuse reflection parameter set to 90%, 100% and 90% respectively, for the three color channels R, G and B, resulting in a light greenish gray. The marble texture on the table is realized with the radiance function "dirt.cal".

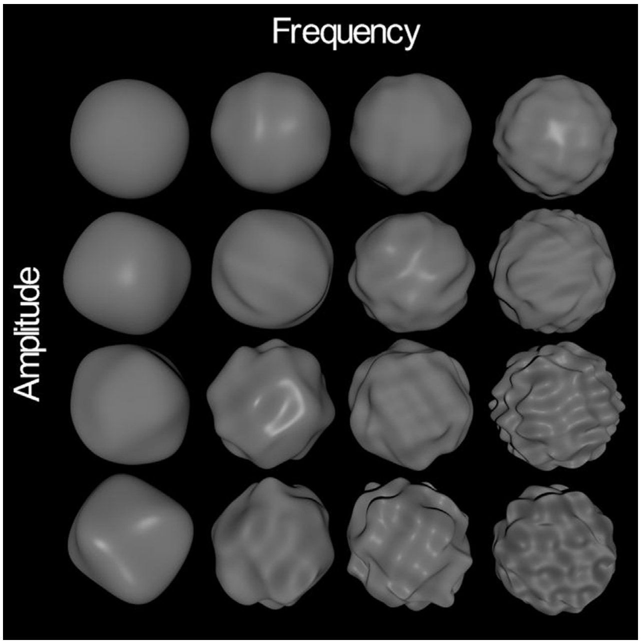

Fig. 4. Blobby Shapes. Sixteen "blobby" objects were generated using low frequency sinusoidal perturbations of spheres. This sinusoidal perturbation was characterized by its sine frequency and amplitude. We generated our shapes with four Amplitude levels and four Frequency levels, in the illustration increasing from top to bottom and from left to right respectively.

frequency, 2) one with high amplitude and low frequency, 3) one with low amplitude and high frequency and 4) one with low amplitude and low frequency. By using different scenes we balanced the combinations of Frequency and Amplitude with Specularity (glossy & matte, M & G in Table 1), so that the four objects defined by the combinations of Frequency were rendered with all the six dispositions of two glossy and two matte objects. The difference between high and low Frequency and high and low Amplitude was also balanced across trials, having all the four combinations: 1) middle low (L−) Frequency vs middle high (H−) Frequency (second and third columns in Fig. 4) and high (H+) Amplitude vs low (L+) Amplitude (first and fourth rows in Fig. 4); 2) middle low (L−) Frequency vs middle high (H−) Frequency (second and third columns in Fig. 4) and middle high (H−) Amplitude vs middle low (L−) Amplitude (second and third rows in Fig. 4); 3) low (L+) Frequency vs high (H+) Frequency (first and fourth columns in Fig. 4) and high (H+) Amplitude vs low (L+) Amplitude (first and fourth rows in Fig. 4); 4) low (L+) Frequency vs high (H+) Frequency (first and fourth columns in Fig. 4) and middle high (H−) Amplitude vs middle low (L−) Amplitude (second and third rows in Fig. 4). The combinations described above are listed in Table 1. The position in the scene was also balanced, having all 24 dispositions, for every combination of the other factors as described above. To balance all the factors, we rendered 576 (24×24) scenes in total. The four reflectances were randomly assigned for every scene, and the objects were placed in the scene with a random rotation on their vertical axis.

Classification analysis on the rendering parameters: We repeated the classification analysis on this new set of shapes to ensure that the general shapes we had chosen had similar properties to the large database we analyzed in the first experiment, i.e. the brightest parts of matte surfaces are good predictors of the surface albedo, whereas they perform relatively poorly when specular reflection is involved.

Analysis: Similar to the initial simulation, for every object in every rendered scene we extracted the luminance distribution of the light coming from the virtual surface; for each of these distributions, we computed the percentile statistics. In every scene, we compared the two matte surfaces and the two glossy surfaces in terms of the diffuse reflection parameter (rendered albedo). Thus for every scene, we had a binary classification of albedo for both the matte and glossy surfaces. We performed our analysis separately for matte and glossy surfaces, using the percentile statistics of each surface to predict its binary value in the binary comparison within each scene (i.e., 0 = darker than the other matte or glossy surface, 1 = lighter than the other surface).

Table 1

Shape Parameters. Every scene contained 4 objects (Object 1–4, first row). For both Frequency and Amplitude, we used 4 levels in our renderings: H + (high), H − (middle high), L − (middle low) and L + (low). In order to balance high Frequency (H+ & H−) and low Frequency (L+ & L−) with high Amplitude (H+ & H−) and low Amplitude (L+ & L−), with Specularity (G & M) we needed 24 scenes.

	Object 1			Object 2			Object 3			Object 4		
Scene	Amp.	Freq.	Spec.	Amp.	Freq.	Spec.	Amp.	Freq.	Spec.	Amp.	Freq.	Spec.
1	H+	H+	M	H+	L+	M	L+	H+	G	L+	L+	G
2	H+	H+	M	H+	L+	G	L+	H+	M	L+	L+	G
3	H+	H+	G	H+	L+	G	L+	H+	M	L+	L+	M
4	H+	H+	G	H+	L+	M	L+	H+	M	L+	L+	G
5	H+	H+	M	H+	L+	G	L+	H+	G	L+	L+	M
6	H+	H+	M	H+	L+	M	L+	H+	G	L+	L+	M
7	H+	H−	M	H+	L−	M	L+	H−	G	L+	L−	G
8	H+	H−	M	H+	L−	G	L+	H−	M	L+	L−	G
9	H+	H−	G	H+	L−	G	L+	H−	M	L+	L−	M
10	H+	H−	G	H+	L−	M	L+	H−	M	L+	L−	G
11	H+	H−	M	H+	L−	G	L+	H−	G	L+	L−	M
12	H+	H−	M	H+	L−	M	L+	H−	G	L+	L−	M
13	H−	H+	M	H−	L+	M	L−	H+	G	L−	L+	G
14	H−	H+	M	H−	L+	G	L−	H+	M	L−	L+	G
15	H−	H+	G	H−	L+	G	L−	H+	M	L−	L+	M
16	H−	H+	G	H−	L+	M	L−	H+	M	L−	L+	G
17	H−	H+	M	H−	L+	G	L−	H+	G	L−	L+	M
18	H−	H+	M	H−	L+	M	L−	H+	G	L−	L+	M
19	H−	H−	M	H−	L−	M	L−	H−	G	L−	L−	G
20	H−	H−	M	H−	L−	G	L−	H−	M	L−	L−	G
21	H−	H−	G	H−	L−	G	L−	H−	M	L−	L−	M
22	H−	H−	G	H−	L−	M	L−	H−	M	L−	L−	G
23	H−	H−	M	H−	L−	G	L−	H−	G	L−	L−	M
24	H−	H−	M	H−	L−	M	L−	H−	G	L−	L−	M

3.2. Results

Fig. 5 represents the performance of the albedo classification on the matte and glossy surfaces, separately conducted on the scenes rendered for the classification experiment. The goal of this analysis is to reveal differences between matte and glossy surfaces, therefore we ignored the shape factors (i.e., Amplitude and Frequency) thus increasing the number of instances that were used to train the classifier. Results are qualitatively similar to the classification results of the initial simulation: for the matte surfaces, performance increases with the percentile; whereas for the glossy surfaces, after approximately the 80th percentile performance decreases with the percentile. The only noticeable exception is that the performance of the highest percentile (the luminance maximum) of the glossy surfaces increases almost to the level of the matte surfaces.

3.3. Discussion

The reason for the discrepancy between the results of the two experiments when it comes to the performance of the luminance maximum is probably due to the fact that the "blobby" shapes we used are all more or less spherical (they cover a large range of surface normals), whereas some of the shapes we initially used (like the cube or the pyramid) only have a small number of faces (as opposite to a broad range of surface normals). Even some of the shapes that are not fully delimited by flat faces, (like the plate, the hand, and many statues and models) have an approximately flat plane on some side. The diagnosticity of each percentile (i.e. the 100th,) is determined both by the average difference in its value between the two reflectance groups (i.e., high and low albedo), as well as by the variability within each distribution. When a surface presents a broad range of normals, it is likely that the most intense illumination point in the light-field is mirrored towards the observer somewhere on the object surface. When a surface only has a few number of faces, the chance that the specular highlight corresponds to the maximum intensity is much smaller. For approximately spherical surfaces (which have a broad range of surface normals), the maximum of the specular reflection component is very stable, corresponding to the maximum of the light-field. The variability added by the specular component to the 100th percentile is minimal, leaving the diffuse component as the main determinant of radiance. We tested this reasoning explicitly with a series of simulations (described in the following paragraph).

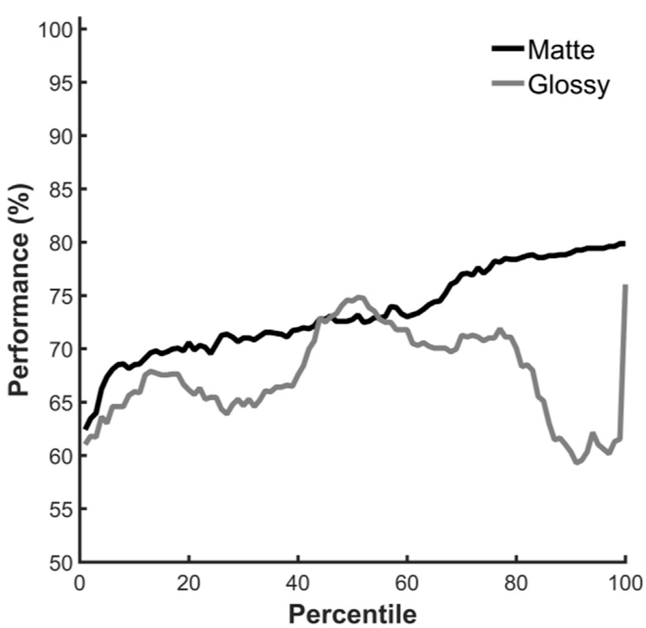

Fig. 5. Classification results for the albedo of the "general shapes". Performance is represented on the y-axis for each percentile (represented on x-axis). The solid lines represent the classification performance for matte (black) and glossy (gray) surfaces. Results are qualitatively similar to the classification results of the natural objects: for the matte surfaces, performance increases with the percentile, whereas for glossy surfaces, after approximately the 80th percentile performance decreases with the percentile. Contrary to the previous results, for the glossy surfaces the performance at the highest percentile (the maximum) increases almost to the level of the performance for the matte surfaces.

Performance highest percentile: We reasoned that the dramatic increase in performance for the maximum (100th percentile) of the glossy surfaces was due to a particularly low variability in the reflection from the specular component for that percentile. We tested this idea by first separately rendering the specular and matte components for the glossy scenes analyzed above. Since classification performance decreases with the distance between the means of the distributions of the two albedo groups (i.e. high and low albedo), and with the variability of these two distributions, we computed and represented standard deviation and mean difference separately (Fig. 6A&B, respectively).

The general shapes used in the renderings for the ranking experiment were divided into low or high albedo, according to the comparison of the diffuse reflectance parameter between the two matte or the two glossy shapes within each scene. Given the two albedo groups, the group average luminances were computed. The difference between these averages is represented in Fig. 6B, for specular and diffuse components individually; and for the renderings where the two components are combined. The luminance of the specular component does not depend on the albedo; therefore, the average difference between the two groups is approximately zero, irrespective of the percentile (black line), whereas for the diffuse component this difference increases with the percentile as we previously found (Toscani et al., 2013, 2015). When the two components were rendered together, the profile of this difference as function of the percentile rank resembles one of the diffuse components alone. It follows that the distance between the averages of the distributions of the groups that the linear classifier was trained to discriminate cannot explain the increase in performance for the highest percentile. In Fig. 6A, the standard deviation of the luminance is represented for each percentile, for the diffuse component alone, for the specular component alone, and for the combined renderings. In the specular component renderings, the highest percentile (maximum) exhibits a particularly low variability (black line) as opposed to the variability in the diffuse component (that increases until approximately the 60th percentile and holds at a similar level until the last percentiles). The reduction in variability in the percentiles extracted from the scenes where the two components were rendered together is thus explained by the reduction in variability in the specular component, which also explains the high performance of the highest percentile in the linear classification.

Range of surface normals: We wanted to test whether this reduction in the variability of the luminance of the last percentile actually depends on the fact that when a surface has a broad range of normals (as opposite to a small number of faces) the highest value is likely to be the brightest illumination point in the light-field. We therefore rendered 400 scenes with one of the "blobby" shapes that we planned to use in the ranking experiment, and 400 with a tetrahedron, the tridimensional shape with the minimum possible number of faces. The object models were placed in the centre of the light-field and randomly rotated on their axis. We used a different light field to test whether the results held independent of the illuminant structure. For the tetrahedron, the highest percentile was highly dependent on the position of the shape in the light-field, and thus quite variable (Fig. 6D), whereas for the "blobby

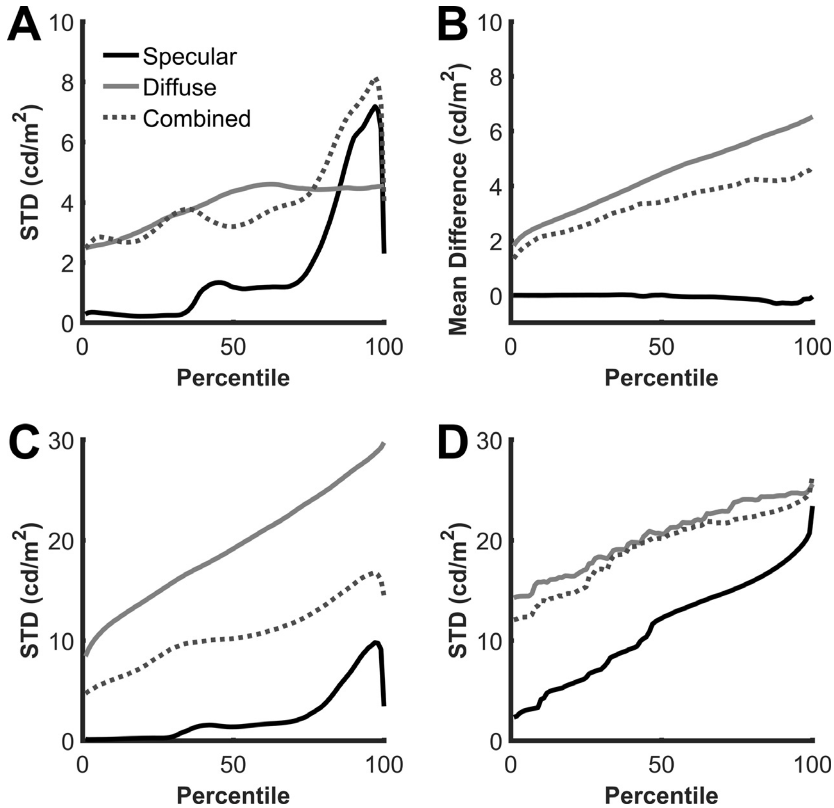

Fig. 6. Variability and effect of albedo on percentile averages for "blobby" shapes. A) Standard deviation (y-axis) of luminance for every percentile (x-axis) for the specular component (black line); the diffuse component (gray line); and the combined (dotted gray line) glossy shapes. B) Difference between luminances associated with the percentiles in the high and low reflectance surfaces, for the specular and diffuse components (individually) and for the combined renderings. C) Standard deviation of luminance for every percentile for the specular component (black line), the diffuse component (gray line), and the combined (dotted gray line) rendering, for one of the general shapes (3rd row 2nd column of Fig. 4) under a light-field different from the one used to render the shapes for the ranking experiment ("Funston Beach at Sunset"). D) Same as in C), but with a tetrahedron. The decreased luminance variability in the highest percentile is completely due to the specular reflection in blobby shapes and is completely absent in the tetrahedron.

shape" it was extremely stable; supporting the idea that the peak in performance in the classification of the glossy surfaces was due to the roundness of the models. In our simulation the illumination fields were roughly characterized by direct lighting from the top left of the observer in addition to a diffuse component (see Fig. 3). Had the angle between the viewing and illumination directions been larger, say with light coming from almost behind the object, the specular highlights would have been shifted to regions with even darker diffuse shading. In this case it is possible that even a very stable maximum would have been poorly informative about lightness, and even surfaces with a broad range of surface normals (as opposite to only a few number of faces) would have yielded the same pattern of result expressed in Fig. 2. Overall, we think it is quite unlikely that the sensitivity of the human visual system is high enough to pick up on the spurious association between the specular highlight intensity and albedo for glossy surfaces. In fact, the association between the maximum of the luminance distribution of glossy surfaces with the observers' judgements is low when compared to the other percentiles (see Fig. 9). This could happen either because observers ignored the specular highlight or because the sensitivity of the visual system is very poor for discriminating these few isolated bright spots when being adapted to a much lower luminance level (Craik, 1938). Furthermore, the results of the experiment where we directly manipulated the luminance bands strongly suggest that observers do not use specular highlights in their lightness judgements. Specifically, the increment or decrement of the last band, coinciding with the specular highlight in the glossy images, had no effect on lightness though these manipulations were clearly visible (see Fig. 11B).

4. Behavioral experiment

We presented our observers with our set of "blobby" shapes and asked them to classify the objects in terms of lightness. We used all the 576 shapes to ensure the same level of generality that we achieved in the simulation.

Observers: Six naïve participants took part in the experiment; all observers had normal or corrected-to-normal visual acuity, and normal color vision. All observers gave written informed consent in agreement with the local ethics committee and in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Perceptual Task: Each observer sat in front of a computer screen and was presented with the 576 rendered scenes. The scenes had a resolution of 1000×1000 pixels and observers set at approximately 62 cm from the centre of the computer monitor. Consequently, the visual scenes subtended 26×26 degrees of visual angle. The task requirement for the observers was to "pick the object which is painted with the lighter gray paint". In each scene the four objects were presented on a virtual marble table (Fig. 3). Observers indicated their choice with the computer mouse by left clicking on the chosen object. The selected object disappeared, leaving three objects in the scene. Again the observer had to pick the object that was painted lighter, this until one single object was left in the scene. Each of the 576 scenes was shown once to every participant. The scenes where presented to each observer in random order.

Analysis: First, we wanted to determine whether the presence of specular highlights affects the accuracy and precision of lightness judgments. Namely, we investigated whether glossy surfaces with the same diffuse component as matte surfaces appear to have the same lightness; or whether the specular component introduces a bias in lightness perception. Additionally, we tested whether the presence of a specular component makes the lightness judgment less precise. Similar to the approach we applied to determine the relative diagnosticity of luminance percentiles with regard to physical albedo, we used linear classification to determine how well each percentile predicts perceived lightness. Observers ranked the four objects in every scene according to their lightness. As a first step we used each individual percentile to classify the ranking of the glossy and matte surfaces separately. We trained the classifier on the whole set of images, leaving out one image at a time.

Ranking positions results: We tested the effect of Specularity on lightness rankings with repeated measures ANOVAs on the observers' average ranks. A 2-way repeated measures ANOVA with Specularity and Amplitude as fixed factors revealed a main effect of Specularity ($F(1,6) = 8.084$, $p < 0.05$), no effect of Amplitude ($F(3,18) = 0.71$, $p = 0.558$), and no interaction ($F(3,18) = 1.719$, $p = 0.199$). A second 2-way repeated measures ANOVA with Specularity and Spatial Frequency as fixed factors revealed a main effect of Specularity ($F(1,6) = 8.084$, $p < 0.05$), no effect of Spatial Frequency ($F(3,18) = 2.96$, $p = 0.06$), and no interaction ($F(3,18) = 2.701$, $p = 0.076$). This results show that glossy surfaces are ranked as darker than matte ones (Fig. 7, A&B).

Ranking precision results: Specular reflection does not depend on the diffuse reflectance component of the surfaces. Therefore, it is not informative about the surface albedo, and specular highlights could be interpreted as luminance noise added to the diffuse component. Since the specular highlights tend to be visible in the proximity of the parts of the surface that receive a more direct illumination (Fleming et al., 2004; Koenderink & van Doorn, 1980); and given that those parts who receive a more direct illumination are the most informative about the surface albedo (Toscani et al., 2013, 2015), it is plausible that lightness perception is less precise for glossy than for matte surfaces. To test this hypothesis, we used the ranking data to measure the precision of lightness perception. In every scene, there were two glossy and two matte surfaces; one rendered with a higher albedo than the other object in each pair. For every pair, each object was ranked consistently or inconsistently with the rendered albedo. For every comparison between the two matte and two glossy surfaces, we computed the probability of the higher albedo surface to be ranked as lighter than the other as a function of the albedo difference between the two surfaces. We considered the steepness of this function as a measure of the precision of the observers. We fitted cumulative Gaussians to the comparison data (example for one observer in Fig. 8A), separately for glossy and matte surfaces. For every observer, we computed Precision as the inverse of the standard deviation of the Gaussians, since the steepness of the cumulative Gaussian function inversely depends on its standard deviation (Fig. 8B).

A repeated measures *t*-test revealed a significant difference in Precision between the rankings of Matte and Gossy surfaces ($t(6) = 5.61$, $p < 0.005$). These experimental results suggest that the visual system is affected by the noise added by the specular component, while judging the lightness of three-dimensional surfaces.

Results for ranking classification: We repeated the classification analysis described before with the difference that binary comparisons were performed on the rankings produced by the observers, instead of on the diffuse reflectance values used to render the scenes. For every scene, the perceived lightness ranking of an object was compared with the lightness ranking of the other with the same Specularity (glossy or matte). These binary comparisons were classified with a binary linear classifier. Classification was done for each individual percentile, separately for glossy and matte surfaces, irrespective of the shape factors (i.e. Frequency and Amplitude). We decided to pool the shape factors together because considering them individually led to noisy and difficult to interpret results. We trained the classifier on the whole set of images, leaving out one image at a time. Fig. 9 represents the classification performance of every percentile of the surfaces' luminance distribution for matte and glossy objects. High performance of the classifier for a specific histogram band indicates that the luminance within that band is highly associated to perceptual judgments.

For matte surfaces, the performance of the classifier tends to increase with the percentiles, with the peak performance for the brightest parts of the images. In contrast, for the glossy surfaces, performance peaks at the darkest parts of the luminance distributions of the objects, and starts to decrease approximately after the 70th percentile. This performance profile qualitatively resembles the one found for the surface albedo: in matte images, the highest parts of the luminance distributions are associated with high performance, whereas in the glossy images, the lowest parts are more

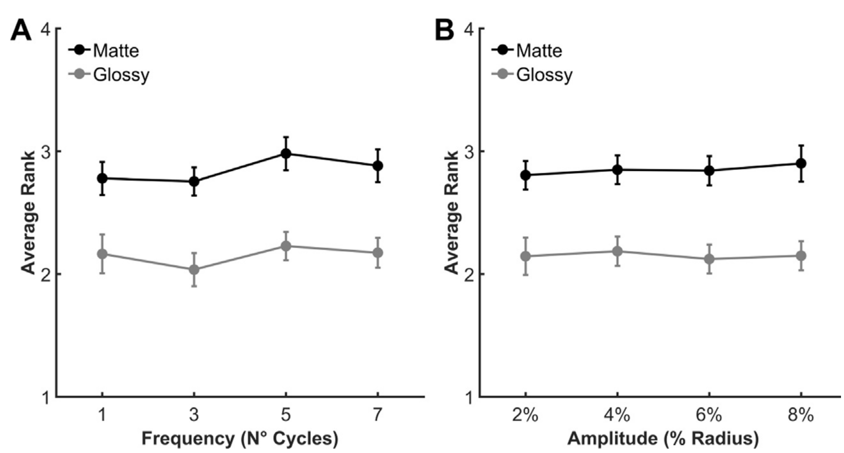

Fig. 7. Ranking results for glossy and matte surfaces. A) Average rankings (y-axis) as a function of stimulus Frequency (x-axis) for glossy and matte surfaces. Black data points represent matte surfaces; gray data points represent glossy surfaces. Error bars represent the standard error of the mean. B) Average rankings (y-axis) as a function of stimulus Amplitude (x-axis) for glossy and matte surfaces. Black data points represent matte surfaces; gray data points represent glossy surfaces. Error bars represent standard errors of the mean. Glossy objects were consistently rated as being darker than matte objects.

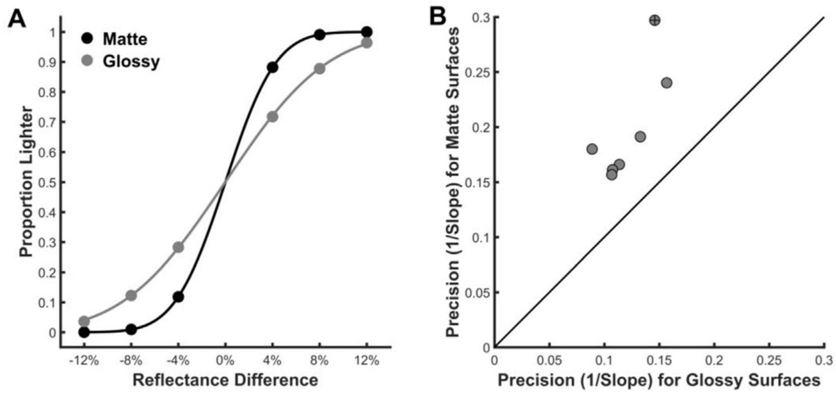

Fig. 8. Precision for glossy and matte surfaces. A) Probability of being ranked as lighter as a function of albedo difference. Example data from observer EK. Cumulative Gaussian functions are fitted separately for glossy and matte surfaces. B) Precision for Matte (y-axis) and Glossy (x-axis) surfaces. Precision is computed as the inverse of the standard deviations of the fitted Gaussian curves. Circles represent the different observers. The circle filled with a cross represents observer EK, whose individual data is shown in A).

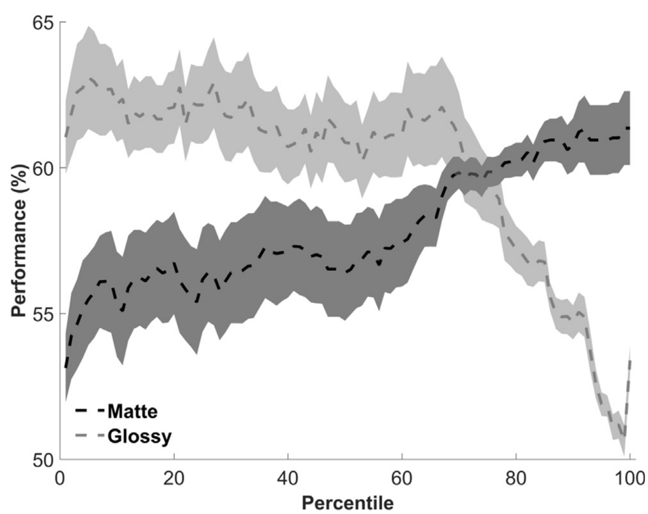

Fig. 9. Classification results for observers' rankings of the "blobby shapes". Performance is represented on the y-axis for each percentile (represented on the x-axis). The dashed lines represent the classification performance for matte (dark gray) and glossy (light gray) surfaces. The shaded areas represent their standard errors.

informative about the surface albedo. This result suggests that observers exploit two different strategies in judging the albedo of matte or glossy surfaces. Namely, for matte surfaces, observers mostly consider the brightest parts of the surfaces' luminance distributions (which are the most informative about their albedos). For glossy surfaces (where the brightest parts tend to correspond to the specular highlights), people tend to base their judgments on lower parts of the luminance distributions.

Results from the classification of human judgments resemble the ones from the classification of surface rendering albedo; however, due to the correlational nature of the experiment, this resemblance could just be explained by the fact that people could do the task with over-chance performance. In order to test this possibility we designed an experiment where we manipulated different parts of the luminance distributions to study how they affect lightness perception.

5. Luminance bands manipulation

In this experiment we selectively manipulated different parts of the luminance histogram of a virtual shape to determine their impact on perceived lightness, for glossy and matte surfaces separately. The previous image analyses showed that the optimal solution is to base lightness judgments of matte surfaces on the brightest parts of their luminance distributions; whereas for glossy surfaces, people should base their judgments on relatively lower luminance distribution bands.

5.1. Methods

Observers: Ten naive observers took part in the experiment. All observers had normal or corrected-to-normal visual acuity. All observers provided written informed consent, in agreement with the local ethics committee, and in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Stimuli: We selected one of the "blobby" shapes with central parameters (Frequency = 5 cycles, Amplitude = 6%) and rendered it with three different random rotations on its vertical axis. The random rotations were balanced across trials. The shape was placed in the same virtual scene used for the ranking experiment (Fig. 10), in the horizontal line closer to the observer, and on the right or left of a sphere. The side position (left or right of the sphere) was balanced across trials. The size of the stimuli was similar to the one in the previous experiment.

Renderings: We again rendered our scenes using RADIANCE software (Ward, 1994) through the MATLAB-based RenderToolbox3 (Lichtman et al., 2007). The "blobby" shapes were rendered in the same manner as the ranking experiment. The matte surfaces were rendered according to a Lambertian reflection model, whereas the glossy surfaces were rendered according to a Ward model (Ward, 1992). The diffuse component was set to 0.4. The specular reflections for the glossy surfaces were defined by the

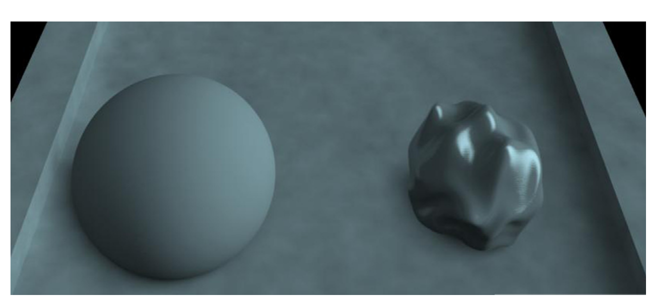

Fig. 10. Visual display used for the luminance bands manipulation experiment.

specular component (set to 0.2), and roughness (set to 0.1). We used "The Uffizi Gallery" light probe as the virtual illuminant for all scenes. The sphere was always totally matte, and was rendered with diffuse component set to 1 (100% reflectance). We then produced images of the sphere (with all other desired diffuse components) by simply multiplying the sphere image for the desired value (comprising values between 0 and 1). This is possible because the sphere is totally matte and convex; therefore, the effect of diffuse reflectance on the reflected light from every point of the surface is linear and there are very few inter-reflections (limited to the areas of the surface facing the table) to potentially introduce nonlinearities.

Bands manipulation: In order to manipulate different parts of the luminance histogram of the "blobby" shapes, we extracted the luminance distribution with a zero reflectance mask (as we did with the other shapes). We then divided it into ten bands according to its deciles, i.e. the first band included all the pixels with a luminance between the minimum value of the luminance distribution and the 10th percentile. In one condition (Increment), the luminance of each band was increased until the mean luminance of the whole distribution was 5% higher; and in another condition (Decrement), it was decreased until the mean luminance of the whole distribution was 5% lower (Fig. 11). The manipulation of the luminance histogram was small enough to avoid causing the stimuli to appear clearly unrealistic and/or to appear of non-uniform albedo, yet it was large enough to affect lightness perception. We interviewed our participants at the end of the experiment and nobody noticed anything unrealistic in the luminance profile of our "blobby" shapes. The luminance manipulation was applied to the 1st, 4th, 7th, and 10th bands. These bands were relatively distant between each other in terms of luminance as compared to the small magnitude of our manipulation. Because of this, only 5.3% of the pixels moved from one band to one of the others that we manipulated; therefore, we are confident that the conclusions we draw from the manipulation of a given histogram portion actually apply to that portion (e.g. when we manipulated the 4th band we could be reasonably confident in drawing conclusions about the middle luminances of our shape). We also obtained baseline lightness matches for the unmodified shapes.

Perceptual Task: Observers had to state whether the sphere or the "blobby" shape was higher in lightness. After every response, the luminance of the sphere was varied following an adaptive staircase method (QUEST; Watson & Pelli, 1983). Separate staircases (and thus Points of Subjective Equality – PSE) were obtained for each manipulated band, for the two conditions, and for the shape in the non-modified version. In order to test observer reliability, we measured two separate PSE values for the non-modified version of the shape. Each staircase terminated after 50 judgments.

5.2. Results

Fig. 12 represents the PSEs for the various bands and for the baseline, in the conditions where the luminance was increased and where it was decreased, for the matte and for the glossy objects (Panel A and B, respectively). The baseline PSEs for matte surfaces are clearly higher than for glossy surfaces, in line with the previous observation that glossy surfaces appear darker than matte surfaces. Qualitatively, for the matte surfaces, the PSEs in the condition where luminance was increased are higher than the PSEs in the condition where the luminance was decreased, though for the third band the average PSEs for the two conditions are very similar. For the glossy surfaces, the PSEs seem different between the two conditions of the luminance increment or decrement, but this difference is highly reduced for the brightest band of the luminance histogram, where the PSEs are basically identical on average.

We first used Bonferroni corrected t-tests to compare the PSEs in the two conditions of luminance increment or decrement for each band. For the matte object, only the manipulation of the brightest band of the luminance histogram caused a significant difference in the PSEs between the two conditions ($ts(9) = 2.5$, 3.44, 0.29 and 6.42; p-values = 0.27, 0.06, 1, 0.001; for the first, second, third and fourth band respectively), whereas only the manipulation of the third band of the glossy object caused a significant difference ($ts(9) = 2.73$, 2.58, 4.72 and -0.65; p-values = 0.185, 0.238, 0.009, 1; for the first, second, third and fourth band respectively). Manipulating the highest luminance band of the glossy objects barely affected the PSE. In a further analysis, we performed for each band a 2-way repeated measures ANOVA; for the first, second, third and fourth band respectively, with direction of manipulation (luminance Increment vs Decrement) and Specularity (matte vs glossy) as factors. The ANOVAs revealed significant interactions for the last two bands ($Fs(1,9) = 0.4651$, 0.0844, 11.3803 and 10.4909; p-values = 0.512, 0.778, 0.008 and 0.01; for the first, second, third and fourth band respectively), indicating that the manipulation affected the glossy object more than the matte one when applied to the third band (60th–70th percentile of the luminance distribution), and the matte surface more than the glossy when applied to the fourth band (90th–100th percentile of the luminance distribution). These results are in accordance with the performance results of the ranking experiment, showing that the most diagnostic percentiles are between the 60th and the 70th percentile for glossy surfaces, whereas the highest percentiles are the most diagnostic for matte surfaces.

6. General discussion

We used extensive physically based lighting simulations to investigate how the different percentiles of the luminance histogram of matte and glossy objects are diagnostic of their surface albedo. Results for glossy and matte surfaces were qualitatively different: the brightest parts were the best predictor of the albedo of the matte surfaces, whereas they poorly predicted the diffuse reflectance component of glossy surfaces. These results demonstrate that relying on the brightest parts of a surface as an estimate of its lightness is a viable strategy only for matte objects, whereas it would be misleading for glossy objects. In two perceptual experiments we tested whether observers judge lightness of glossy and matte surfaces differently. Glossy surfaces appeared darker than matte ones (see Figs. 8 and 12), and observers were less precise in judging their lightness. This result is consistent with the idea that human observers ignore the specular reflection component when judging lightness of glossy surfaces. We also used the percentiles of the luminance distributions of the objects to predict observers' judgments. Classification results showed that the highest percentiles highly correlated with human judgments of matte surfaces and poorly correlated with judgments of glossy surface. The percentiles within the darker half of an object's luminance distribution where the best predictors of the lightness rankings in the case of glossy surfaces. In order to test whether this association reveals a different reliance of the observer on surface areas of different brightness, we independently manipulated different sectors of the virtual objects' luminance distribution. Changes in the brightest parts prominently affected the lightness judgments in the case of matte objects, whereas they had no impact on the lightness of glossy objects.

6.1. Simulation

In our first lighting simulation, we used a large set of three-dimensional models under different illuminations to study which

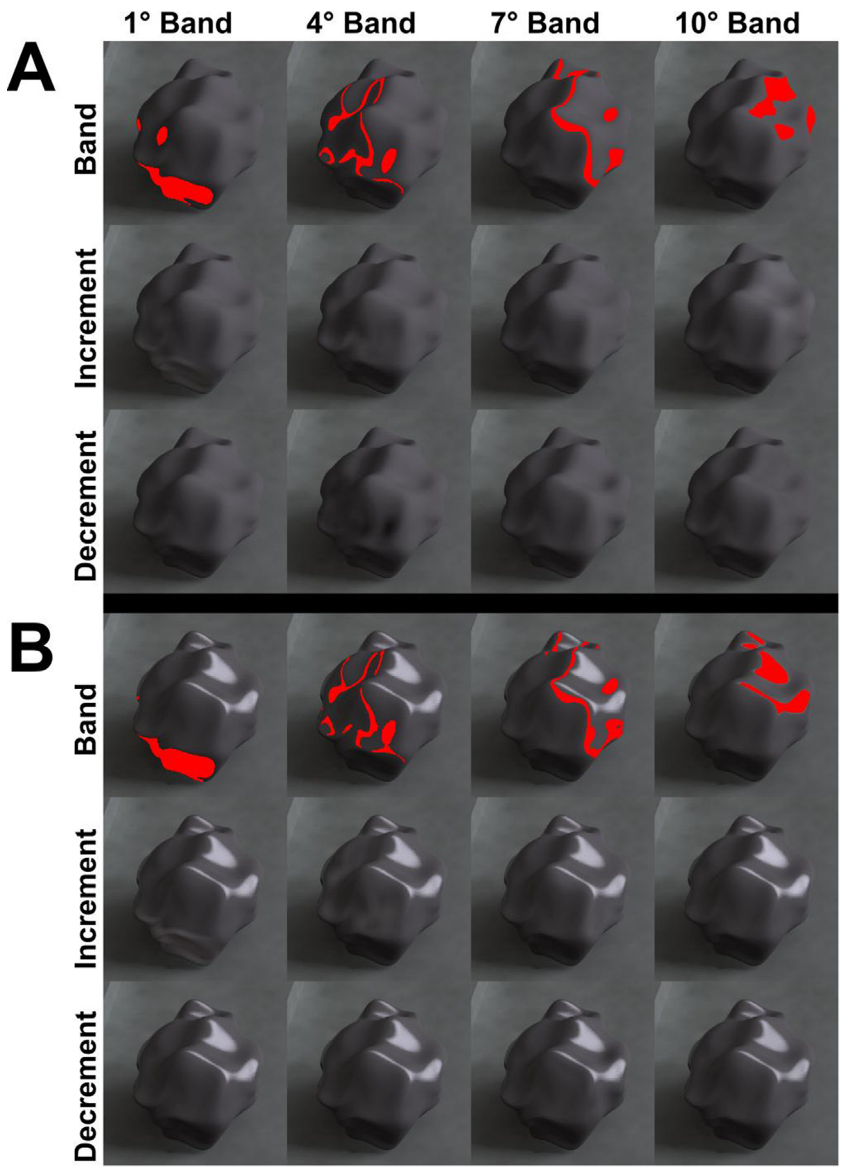

Fig. 11. Luminance bands manipulation. Only the bands used in the experiment are shown (1st, 4th, 7th and 10th bands). The top row represents in red the pixels of the shape that belong to a certain histogram band, from the 1st to the 10th, from left to right. The second row represents the experimental stimuli for the manipulation where the luminance was increased, and the third row represents the stimuli for the manipulation where the luminance was decreased. A) & B) panels describe the manipulation for the matte and the glossy shapes, respectively.

parts of the luminance distributions of the rendered surfaces are particularly informative about the surfaces' albedo. The brightest parts of the luminance distributions of matte surfaces performed best in classifying their diffuse component. This result is analogous to what we previously found with a more limited set of shapes and under one single illumination (Toscani et al., 2013, 2015). On the other hand, the brightest parts of the luminance distributions of the glossy objects performed relatively poorly. One aspect that almost all models of reflection share (Fleming, 2014) is the distinction between a diffuse and specular reflection components; the first one scattered in all directions, and the second one in a single direction according to the law of Hero (e.g. Barrow & Tenenbaum, 1978; Giesel & Gegenfurtner, 2010; Klinker, Shafer, & Kanade, 1990; Shafer, 1984).The model we used for our simulation implements the two components separately, and we classified the rendered objects according to their diffuse component. Since specular highlights tend to be much brighter than diffusely reflected light, because the reflection is more focused (Thompson et al., 2011) they affect the highest percentile of the luminance distributions of the glossy surfaces, without being informative about its diffuse reflection component. Thus the specular highlights can be considered luminance noise added to the information about the surface diffuse component. In a perfectly matte surface, the brightest parts of the luminance distribution are the most informative about reflectance. Whether the noise added by the specular reflection is enough to make those parts poorly informative about the diffuse reflection component relative to the rest of the distribution is an empirical question. Our results show that specular reflection actually makes the brightest part of the luminance distributions of glossy surfaces poor predictors of their diffuse reflection components.

6.2. Glossy shapes appear darker

Ranking results (and also the results of the band manipulation experiment, as it is shown by the PSEs for the baselines) show that

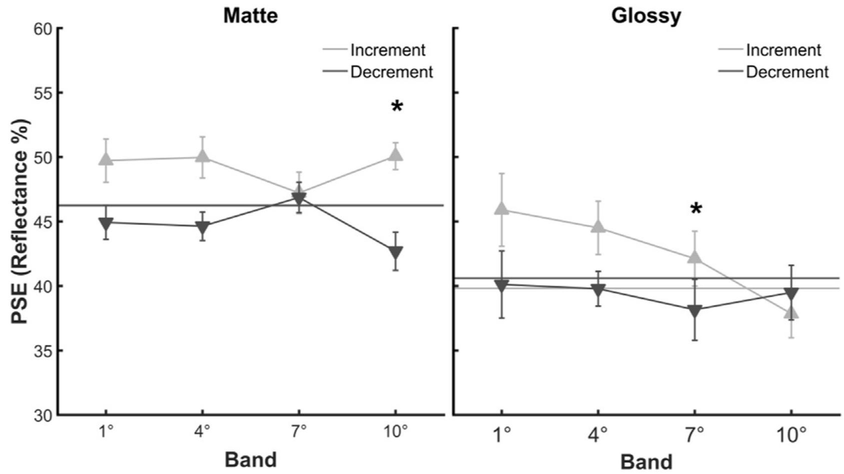

Fig. 12. PSE for the manipulated matte and glossy objects, in the different bands of the luminance histograms. The y-axis represents the PSEs obtained by the QUEST algorithm. The x-axis represents the band of the luminance histogram affected by the luminance increment (light gray lines and upward triangles) or decrement (dark gray and downward triangles). The horizontal straight lines represent the PSEs for the baseline condition, where the luminance of the image was not changed-the baselines were measured with two staircases. Colors representing the baselines are arbitrary. Stars indicate significant PSE differences between the luminance increment and decrement conditions. A and B panels represent the PSEs for the matte and glossy objects, respectively.

glossy objects tend to appear darker than matte objects. This finding is compatible with the idea that lightness is computed by the visual system from the luminance distribution of glossy surfaces (excluding the light coming from the specular highlights). One possible reason for this is that specular highlights tend to appear in the proximity of the maxima of the diffuse component (Fleming et al., 2004; Koenderink & van Doorn, 1980), effectively masking them for the purpose of recovering diffuse reflectance. Another reason might lie in the physics of reflection. It is indeed difficult to equate the physical diffuse reflectance of a matte and glossy surface. The difficulty arises because the incident light reflected by the specular component cannot also be reflected by the diffuse component and vice versa. In our rendering we defined the specular and the diffuse components according to the Ward model (Ward, 1992), where diffuse reflections are proportional to a component of incident illumination that discounts the illumination lost through specular reflection. According to this definition, glossy surfaces physically reflect less light towards the observer compared to matte surfaces with the same diffuse reflection component. We have reasons to believe that this difference in reflected light is in principle large enough to explain our finding that glossy surfaces are perceived as darker than matte surfaces. As a first step we estimated the difference in perceived albedo between glossy and matte surfaces from the luminance bands manipulation experiment. We did this by subtracting the PSE of the non-manipulated glossy surfaces (40% – see baselines in Fig. 12) from the PSE of the non-manipulated matte surfaces (46%). The perceptual effect in our data was thus a 6% reduction diffuse reflectance. As a second step we computed the difference image between the glossy and the matte image for each of the renderings used in the luminance bands manipulation experiment. We divided the luminance histograms of these difference images in fifty bands and selected the band where the negative difference was larger. As a third step we rendered, for each of the original shapes, 50 equivalent matte shapes with a lower diffuse component (according to the Ward model), and subtracted from them the corresponding original matte shape rendered with the original albedo. We then selected the diffuse reflectance level for which the average luminance in the pixels corresponding to the sub-band selected at step 2 was the closest to the one of the glossy image. This procedure gave us a measure of how darker in albedo a matte surface needs to be in order to match the maximum luminance decrease introduced by gloss. This estimate is again 6%, approximately the same as the difference in appearance we observed in the luminance band manipulation experiment. Of course it is unlikely that observers perfectly isolate the band which shows the largest luminance reduction, so we cannot exclude that perceptual factors also play a role. For instance, black glossy surfaces appear shinier than white glossy surfaces despite having identical specular reflectance (Billmeyer & O'Donnel, 1987). This effect might be mediated by a contrast effect, whereby the specular highlights appear brighter being surrounded by an overall darker shape. Similarly, the overall surface of an object might be perceived to be darker when the bright highlights are added. We do not have enough evidence at the moment to tell apart these possibilities and we hope that future studies will shine some light on this issue.

6.3. Specular highlights are discounted

We confirmed the idea that the brightest parts of the luminance distributions of glossy surfaces have a poor impact on lightness perception with an experiment where we manipulated different parts of the luminance histograms of matte and glossy surfaces. A luminance increment or decrement of the highest luminance parts clearly affected lightness perception of matte surfaces, but had no effect on glossy objects, suggesting again that specular highlights are discounted in lightness perception. This idea is consistent with results of Beck (1964). He measured perceived lightness of matte and glossy flat isolated real surfaces. The glossy surfaces presented a vertically elongated visible specular highlight. When the means of the matte and the glossy surfaces were equalized, the glossy surfaces appeared darker, but when all surfaces except the region where the highlight was visible were equalized in luminance, there was no longer a difference in lightness perception. These results are also consistent with the idea that the visual system computes lightness from the luminance distributions of glossy surfaces after excluding the light coming from the specular highlights. The fact that the visual system excludes specular reflection from the computations used to perceive lightness is not trivial, since in many studies it has been shown that lightness perception is affected by brightness variations on the surface (Ripamonti et al., 2004; Robilotto & Zaidi, 2004; Toscani et al., 2015; Zdravković, 2008).

6.4. Lightness predictors

In our study, we aimed to predict lightness judgments with the percentiles of the luminance distributions of our "blobby" shapes. Since our shapes were uniform in paint, with no texture, the luminance variations across their surfaces tended to be smooth and the different percentiles basically represented different areas on the surface. This means that our predictors were directly accessible to the visual system as the luminance at some location in the input image. In contrast, other works proposed that the visual system estimates the surface albedo through other image statistics which do not necessarily represent a luminance value on the surface (for instance, the skewness of the luminance distribution Motoyoshi et al., 2007; Sharan et al., 2008). The skewness of the luminance histogram of our "blobby" shapes could predict observers' judgments in the ranking experiment better than at chance level (59% of correct classification for both matte and glossy objects tested separately). A logistic regression revealed a negative correlation between lightness rankings and skewness (mean $\beta = -0.3114$ for matte images and mean $\beta = -1.7301$), which reached statistical significance in a between-subjects *t*-test only for the glossy surfaces ($t6 = -8.1557$, $p < 0.001$). Glossy surfaces tend to have positively skewed intensity histograms because only small regions are very bright due to specular highlights (Thompson et al., 2011). The strategy of excluding the specular reflection from the lightness computation could explain the negative correlation between skewness and perceived lightness in the glossy surfaces. It is to notice that complex factors also have an effect on lightness perception, such as object shape (Knill & Kersten, 1991; Koffka, 1935), scene geometry (e.g. Gilchrist, 1980; Radonjić, Todorović, & Gilchrist, 2010), interpretation of transparent surfaces and junctions (e.g. Adelson, 1993, 2000; Anderson, 1997, 2003; Anderson & Winawer, 2005, 2008; Metelli, 1970, 1974a, 1974b). These effects require interpretation of the scene and are not explained by simple lateral inhibition between retinal neurons, which filters out shallow intensity gradients, owing mostly to illumination effects. More generally, perceived surface reflectance properties depend on several cues about the surface geometry, such as its two-dimensional profile, texture and motion (Marlow & Anderson, 2016; Marlow, Todorović, & Anderson, 2015) and some of these spatial aspects which cannot be represented in a luminance histogram, are crucial for the perception of reflectance properties of a surface (Anderson & Kim, 2009; Kim, Marlow, & Anderson, 2011). These spatial factors are neglected in the present work, where we focused on the histogram bands of the luminance distributions of our matte and glossy virtual surfaces.

While we suggest that specular highlights are discounted for the purpose of lightness estimation, a theory of how specular highlights are identified is beyond the scope of the present work. However, specular highlights tend share the same orientation as the diffuse shading that surrounds them (Anderson & Kim, 2009; Beck & Prazdny, 1981; Todd, Norman, & Mingolla, 2004) and they tend to appear in regions near (but not coincident) the luminance maxima in diffuse shading (Fleming et al., 2004; Koenderink & van Doorn, 1980). These constraints can be used to discriminate specular reflections which tend to appear as matte when these constraints are violated (Beck & Prazdny, 1981; Kim et al., 2011; Marlow, Kim, & Anderson, 2011). These cues are likely to be used also to detect specular highlights in our scenes. It is worth to mention that in dynamic scenes the visual system could detect specular highlight by exploring their motion characteristics (Doerschner, Kersten, & Schrater, 2011) or their position in depth (Blake & Bülthoff, 1990). Concerning our results, it is yet to be investigated whether observers discount specular reflections in a similar fashion at different levels of glossiness. For instance, it is hard to tell whether observers would behave the same way when specularity is barely visible as they did when confronted with our stimuli (which clearly express specular highlights). Additionally, it is to notice that we rendered specular highlights according to the Ward model simulating the appearance of plastic. In other materials such as metal or glass, specular reflection is characterized by different physical properties which might lead to different perceptual strategies.

7. Conclusions

The present results overall extend our previous finding that human observers apply the optimal strategy in detecting the reflectance of matte surfaces. They rely mostly on the brightest parts of the object's surface, which are also the most informative. Here we show that observers discard the highlights (which are less informative), and rely on relatively darker segments of the luminance distribution (which are more informative about reflectance) when confronted with glossy objects. While this strategy is optimal for the purpose of discriminating between the relative lightness of glossy objects, it comes at a cost. The observers' inability to interpolate the luminance profile masked by the highlights impairs lightness constancy, whereby glossy surfaces appear darker than matte ones. This might be a general sign that humans are more adapted or trained to compare the visual quality of materials belonging to the same class, whereas being able to make absolute comparisons between materials belonging to different classes is a less relevant ability.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft grant DFG GE 879/9 and by BMBF grant 01GQ1304. Lorelei Alley commented on a previous draft of this manuscript and helped us enormously to improve grammar and style. We specially thank Phillip Marlow for his constructive, extensive and expert contribution as a peer reviewer. We are also very thankful to Steven Cholewiak who shared with us his blobby shapes generation code.

References

Adams, W. J., Kerrigan, I. S., & Graf, E. W. (2016). Touch influences perceived gloss. *Scientific Reports, 6*.
Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science, 262*(5142), 2042–2044.
Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neuroscience* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press.
Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception, 26*(4), 419–453.
Anderson, B. L. (2003). The role of occlusion in the perception of depth, lightness, and opacity. *Psychological Review, 110*(4), 785–801.
Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision, 9*(11), 1–17.
Anderson, B. L., & Winawer, J. (2005). Image segmentation and lightness perception. *Nature, 434*(7029), 79–83.
Anderson, B. L., & Winawer, J. (2008). Layered image representations and the computation of surface lightness. *Journal of Vision, 8*(7), 1–22.
Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics, 54*(4), 446–456.
Attewell, D., & Baddeley, R. J. (2007). The distribution of reflectance's within the visual environment. *Vision Research, 47*(4), 548–554.
Barrow, H., & Tenenbaum, J. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds.), *Computer vision systems* (pp. 3–26). New York: Academic Press.
Beck, J. (1964). The effect of gloss on perceived lightness. *The American Journal of Psychology, 77*(1), 54–63.
Beck, J., & Prazdny, S. (1981). Highlights and the perception of glossiness. *Attention, Perception, & Psychophysics, 30*(4), 407–410.
Billmeyer, F. W., & O'Donnel, F. X. (1987). Visual gloss scaling and multidimensional scaling analysis of painted specimens. *Color Research & Application, 12*(6), 315–326.

Blake, A., & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature, 343*(6254), 165–168.
Cholewiak, S. A., & Fleming, R. W. (2013). Towards a unified explanation of shape from shading and texture. *Journal of Vision, 13*(9). 258–258.
Cholewiak, S. A., Kunsberg, B., Zucker, S., & Fleming, R. W. (2014). Predicting 3D shape perception from shading and texture flows. *Journal of Vision, 14*(10). 1113–1113.
Cholewiak, S., Vergne, R., Kunsberg, B., Zucker, S., & Fleming, R. (2015). Distinguishing between texture and shading flows for 3D shape estimation. *Journal of Vision, 15*(12). 965–965.
Cornsweet, T. (1970). *Vision perception*. New York: Academic Press.
Craik, K. J. W. (1938). The effect of adaptation on differential brightness discrimination. *The Journal of Physiology, 92*(4), 406–421.
Debevec, P. (2008). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography (p. 32). In Presented at the ACM SIGGRAPH 2008 classes, ACM.
Doerschner, K., Kersten, D., & Schrater, P. R. (2011). Rapid classification of specular and diffuse reflection from image velocities. *Pattern Recognition, 44*(9), 1874–1884.
Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research, 94*, 62–75.
Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision, 4*(9), 798–820.
Gelb, A. (1929). Die "Farbenkonstanz "der Sehdinge. Springer. In W. A. von Bethe (Ed.), *Handbuch normalen und pathologischen Psychologie* (pp. 594–678). Berlin: Springer-Verlag.
Giesel, M., & Gegenfurtner, K. R. (2010). Color appearance of real objects varying in material, hue, and shape. *Journal of Vision, 10*(9), 1–21.
Gilchrist, A. L. (1980). When does perceived lightness depend on perceived spatial arrangement? *Perception & Psychophysics, 28*(6), 527–538.
Heasly, B. S., Cottaris, N. P., Lichtman, D. P., Xiao, B., & Brainard, D. H. (2014). RenderToolbox3: MATLAB tools that facilitate physically based stimulus rendering for vision research. *Journal of Vision, 14*(2), 1–22.
Heath, T. L. (1921). *A history of Greek mathematics (Vol. 1)*. Clarendon.
Kim, J., Marlow, P., & Anderson, B. L. (2011). The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision, 11*(9), 1–19.
Kim, J., Marlow, P. J., & Anderson, B. L. (2012). The dark side of gloss. *Nature Neuroscience, 15*(11), 1590–1595.
Klinker, G. J., Shafer, S. A., & Kanade, T. (1990). A physical approach to color image understanding. *International Journal of Computer Vision, 4*(1), 7–38.
Knill, D. C., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature, 351*(6323), 228–230.
Koenderink, J. J., & van Doorn, A. J. (1980). Photometric invariants related to solid shape. *Journal of Modern Optics, 27*(7), 981–996.
Koffka, K. (1935). *Gestalt psychology*. NY: Harcourt, Brace & World.
Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *JOSA, 61*(1), 1–11.
Lichtman, D. P., Xiao, B., & Brainard, D. H. (2007). RenderToolbox: A MATLAB Toolkit for Hyperspectral Rendering with Radiance and PBRT (Vol. October 1–2). Presented at the 6th International Radiance and HDR Scientific Workshop, Twin Cities, Minnesota.
Maloney, L. T., & Brainard, D. H. (2010). Color and material perception: Achievements and challenges. *Journal of Vision, 10*(9), 1–6.
Marlow, P. J., & Anderson, B. L. (2016). Motion and texture shape cues modulate perceived material properties. *Journal of Vision, 16*(1), 1–14.
Marlow, P., Kim, J., & Anderson, B. (2011). The role of brightness and orientation congruence in the perception of surface gloss. *Journal of Vision, 11*(9), 1–12.
Marlow, P. J., Todorović, D., & Anderson, B. L. (2015). Coupled computations of three-dimensional shape and material. *Current Biology, 25*(6), R221–R222.
Marr, D. (1982). *Vision*. San Francisco: Freeman.
Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics, 13*, 59–66.
Metelli, F. (1974b). The perception of transparency. *Scientific American, 230*, 90–98.
Metelli, F. (1974a). 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.
Motoyoshi, I., Nishida, S. 'ya, Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447*(7141), 206–209.
Muryy, A. A., Fleming, R. W., & Welchman, A. E. (2016). "Proto-rivalry": How the binocular brain identifies gloss. *Proceedings of the Royal Society B, 283*(1830), 20160383.
Muryy, A. A., Welchman, A. E., Blake, A., & Fleming, R. W. (2013). Specular reflections and the estimation of shape from binocular disparity. *Proceedings of the National Academy of Sciences, 110*(6), 2413–2418.
Nishida, S. Y., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *JOSA A, 15*(12), 2951–2965.
Norman, J. F., Todd, J. T., & Orban, G. A. (2004). Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science, 15*(8), 565–570.
Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of Vision, 10*(9), 1–19.
Olkkonen, M., & Brainard, D. H. (2011). Joint effects of illumination geometry and object shape in the perception of surface reflectance. *i-Perception, 2*(9), 1014–1034.
Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research, 41*(24), 3145–3161.
Poggio, T., & Koch, C. (1985). Ill-posed problems in early vision: From computational theory to analogue networks. *Proceedings of the Royal Society of London: Series B, Biological Sciences, 226*(1244), 303–323.
Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*(6035), 314–319.
Radonjić, A., Todorović, D., & Gilchrist, A. (2010). Adjacency and surrounded ness in the depth effect on lightness. *Journal of Vision, 10*(9), 1–16.
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), 747–763.
Robilotto, R., & Zaidi, Q. (2004). Limits of lightness identification for real objects under natural viewing conditions. *Journal of Vision, 4*(9), 779–797.
Schneider, N., & von Campenhausen, C. (1998). Color and lightness constancy in different perceptual tasks. *Biological Cybernetics, 79*(6), 445–455.
Shafer, S.A. (1984). Using Color to Separate Reflection Components. DTIC Document.
Sharan, L., Li, Y., Motoyoshi, I., Nishida, S. 'ya, & Adelson, E. H. (2008). Image statistics for surface reflectance perception. *JOSA A, 25*(4), 846–865.
Thompson, W., Fleming, R., Creem-Regehr, S., & Stefanucci, J. K. (2011). *Visual perception from a computer graphics perspective*. CRC Press.
Todd, J. T., Norman, J. F., & Mingolla, E. (2004). Lightness constancy in the presence of specular highlights. *Psychological Science, 15*(1), 33–39.
Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2015). Effect of fixation positions on perception of lightness (p. 93940R–93940R). Presented at the IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics.
Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2013). Optimal sampling of visual information for lightness judgments. *Proceedings of the National Academy of Sciences, 110*(27), 11163–11168.
Ward, G. J. (1992). Measuring and modeling anisotropic reflection. *ACM SIGGRAPH Computer Graphics, 26*(2), 265–272.
Ward, G.J. (1994). The RADIANCE lighting simulation and rendering system (pp. 459–472). In Presented at the Proceedings of the 21st annual conference on Computer graphics and interactive techniques, ACM.
Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*(2), 113–120.
Wiebel, C. B., Toscani, M., & Gegenfurtner, K. R. (2015). Statistical correlates of perceived gloss in natural images. *Vision Research, 115*, 175–187.
Zdravković, S. (2008). Lightness constancy: Object identity and temporal integration. *Psihologija, 41*(1), 5–20.