



OPEN

Scale-invariance in brightness illusions implicates object-level visual processing

SUBJECT AREAS:

NEUROSCIENCE
PSYCHOLOGY
VISUAL SYSTEM
OBJECT VISIONErica Dixon¹, Arthur Shapiro¹ & Zhong-Lin Lu²¹Department of Psychology American University, Washington, DC, USA, ²Ohio State University, Columbus, OH, USA.Received
1 July 2013Accepted
20 December 2013Published
29 January 2014Correspondence and
requests for materials
should be addressed to
A.S. (arthur.shapiro@
american.edu)

Brightness illusions demonstrate that an object's perceived brightness depends on its visual context, leading to theoretical explanations ranging from simple lateral inhibition to those based on the influence of knowledge of and experience with the world. We measure the relative brightness of mid-luminance test disks embedded in gray-scale images, and show that rankings of test disk brightness are independent of viewing distance, implying that the rankings depend on the physical object size, not the size of disks subtended on the retina. A single filter that removes low spatial frequency content, adjusted to the diameters of the test disks, can account for the relative brightness of the disks. We note that the removal of low spatial frequency content is a principle common to many different approaches to brightness/lightness phenomena; furthermore, object-size representations--as opposed to retinal-size representations--inherently remove low spatial frequency content, therefore, any process that creates object representations should also produce brightness illusions.

Simultaneous brightness contrast (SBC) is a visual phenomenon in which a mid-luminance test patch appears brighter when placed against a black background and darker when placed against a white background. SBC shows a deviation between what we perceive (brightness) and an objective measure of the environment such as a luminance reading from a photometer; the phenomenon therefore allows insight into how neural processing in the retina and in the brain shapes our perceptual world. The most common historical explanation for SBC has been lateral inhibition, a process in which signals generated in response to the surrounding field inhibit the strength of signals generated by responses to the central test patch. Lateral inhibition is considered a “low-level,” bottom-up approach to understanding SBC because it is the wiring in the retina that shapes the visual response to the image and produces the difference between perceived brightness and physical luminance levels; this approach does not require any higher level response mediated by knowledge or memory. While there have consistently been other theories concerning brightness and lightness^{1,2} over the past 20 years, numerous researchers have produced displays and configurations aimed at arguing that lateral inhibition (and other low-level explanations) are inadequate for explaining SBC-type phenomena. The alternative theories propose that our perception of brightness is influenced and mediated by our previous knowledge about the world, or by the conceptual frameworks in which the images are placed³⁻⁵.

Recently, Gilchrist and Radonjić⁴ developed a powerful technique in which observers report on the appearance of identical mid-luminance test spots placed in the context of natural gray-scale images. The appearance of the test patches did not depend upon the test spot's immediate spatial context, but instead seemed to depend more upon the test spots' illumination framework (i.e., how the spot was organized relative to lighting and surfaces in the scene). The authors therefore concluded that the appearance of the test patches “[could not] be explained without an explicit representation of the structure of illumination in the scene” and stated that they “are aware of no low-level [i.e., lateral inhibition] approach that can account for our obtained pattern of results.”

Using similar types of displays, however, Shapiro and Lu⁶ found that the relative rankings of the test patches could be accounted for simply by filtering the low spatial frequency content from an image. For example, Figure 1 shows test disks on a natural scene image (a) and the same image post-filtering (b); in 1a, all the disks have a pixel value of 128, but in 1b, the values of the disks mimic the perception of brightness (i.e., disk E which is perceived as being darkest now has the lowest pixel value (57)). To remove low spatial frequency content, Shapiro and Lu used a filter that followed the following equation:

$$\text{New Image} = \text{Original Image} - \text{Original Image} * H + \text{Constant} \quad (1)$$

where H is an averaging kernel of diameter N pixels that blurs the image. The basic idea, then, is that the filter subtracts a blurred image from the original image and a constant is added to bring the image values back into a

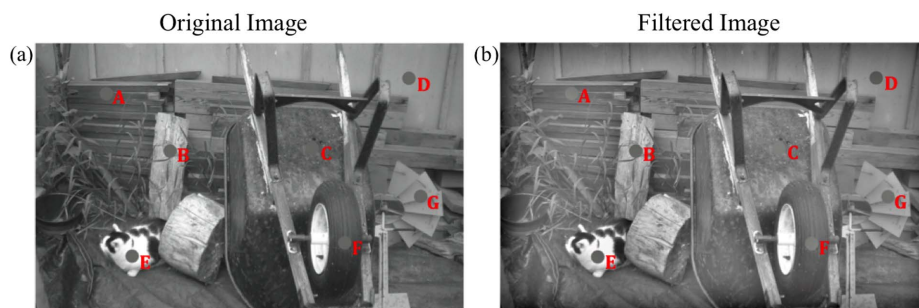


Figure 1 | Test disks placed in grayscale image (Image 1). (a) Original (i.e. unfiltered) image with seven test disks labeled A–G; each disk has a pixel value of 127. (b) Image in (a) filtered to remove low spatial frequency content; test disks in image (b) have physically different pixel values (disk A has a pixel value of 155, B = 95, C = 152, D = 102, E = 57, F = 165, G = 112). All images were taken by Erica L. Dixon.

viewable range. There are many equivalent ways of constructing such a filter, but the simplest is to use the Adobe Photoshop high pass filter function, which allows the observer to control the amount of low spatial frequency content removed from an image. The image in Figure 1b is constructed with the Photoshop filter.

Initially, it might appear that equation (1) is just a new way to express lateral inhibition. After all, equation (1) and lateral inhibition are both types of high pass filters. Indeed, if equation (1) had a fixed filter size, then the equation would be very similar to models that propose that lateral inhibition extends over a larger spatial range^{7,8}. Similarly, it might appear that equation 1 is a method for discounting the illuminant, since models that remove shadows or attempt to calculate surface reflectance also make use of similar types of high-pass filters^{3,9}.

However, the Shapiro and Lu filter embodies a broader theory about how object perception relates to surface features, such as color and brightness. The brain represents the world in terms of objects instead of pixels of light^{10–12}. Object representation is its own form of spatial filter, since object representations do not require the encoding of low spatial frequency content. Consider, for example, a coffee mug with a logo placed on a table. Theories of object coding^{10,12,13} propose that the visual representation of the mug contains the mug's associated features; such a representation would necessarily filter out low spatial frequency content because the mug's important information (the logo design, reflectance patterns, etc.) is carried within its boundaries. Information about the illumination and shadows is carried in the low spatial frequency range, and as far as object perception is concerned, is unnecessary and unwanted; part of the goal of object representation therefore would be to remove the low spatial frequency information.

To be clear, we are not suggesting that the object representations are the only spatial filters in the visual system; clearly, spatial filtering occurs during many stages of visual processing – for example, eye movements seem to create very early and adaptable spatial filters in retinal ganglion cells¹⁴. We are suggesting, however, that one of the end results of visual processing is a perceptual world parsed into objects and in which color and brightness are bound to those objects; we argue that information in this type of representation is appropriately filtered to produce brightness illusions. The Shapiro and Lu hypothesis, then, is that the spatial filter for brightness perception depends upon the size of an object, i.e., the physical or measurable size of the object within a scene or image, as opposed to simple lateral inhibition or a multi-scale filter. This hypothesis is consistent with reports from the spatial vision literature that suggest that some aspects of visual encoding can be best expressed in terms of the size of the object relative to the whole image (i.e., object frequency) rather than the size of the test disks on the retina (i.e., retinal spatial frequency); such a system has been suggested for the detection of letters in noise¹⁵. Here we examine whether the filter in the Shapiro and Lu⁶ model can be best accounted for in terms of relative object-size in the

image or in terms of the retinal projection of the object. As with Shapiro and Lu, observers rank the brightness of test disks placed within natural images, but here the disks are of multiple sizes and the images are viewed at a range of distances (see Figure 2).

As a general rule, we find that the test patches do not change their relative brightness as a function of viewing distance, a result that is consistent with other reports of scale invariance and brightness¹⁵, but we are the first to have tested this principle in the Gilchrist and Radonjić natural image paradigm (we note in the discussion some of the important consequences of using natural images for brightness perception). Our results cannot be accounted for by simple lateral inhibition, in which the size of the filter remains constant, but could be accounted for by adjusting the size of the filter kernel in equation (1); to work, the filter kernel must be adjusted to about the size of the object. Our results could also be accounted for by several different classes of models: a) adjusting the response of multiple filters based on the output of the most active spatial frequency channel^{16,17} (i.e., multi-scale model); b) a higher-level inferential process that discounts the illumination (inference-based model); and c) relative ranking of the luminance levels of frameworks within a scene (i.e., anchoring theory). All of these approaches have their proponents and detractors, and we do not intend our very simple filter to replace these approaches; instead, we suggest that a high-pass filter tuned to object-size can be considered a common principle shared by multi-scale, inference-based, and gestalt approaches to brightness/lightness perception.

Results

Observer rankings of the brightness of test disks placed in natural images can be matched by a filter that removes low spatial frequency content from the image⁶. Here we ask if the cut-off frequency of the filter depends upon the absolute size of test disks and/or on the distance at which the image is viewed. If the filter depends upon the size of the object as projected onto the retina, then we would expect the filter's optimal range of spatial frequency response to change as a function of both the viewing distance and the size of the disks. If the filter depends on the absolute size of the object, measured here as diameter in pixels, then we would expect the opposite to happen: the optimal frequency would change as a function of the size of the object, but not as a function of viewing distance.

Observer rankings of brightness as a function of distance and test disk size. The first and most obvious empirical finding is that for a fixed object size, observer rankings stay constant as function of viewing distance. For example, Figure 3 shows the average rankings for Image 1. The lines in each panel (a–d) represent the ranking of a disk of a fixed pixel size at all viewing distances (120 pixels excluded for space). The average rankings are remarkably consistent across all distances, meaning that the relative brightness rankings did not change as the observer moved away from the image.

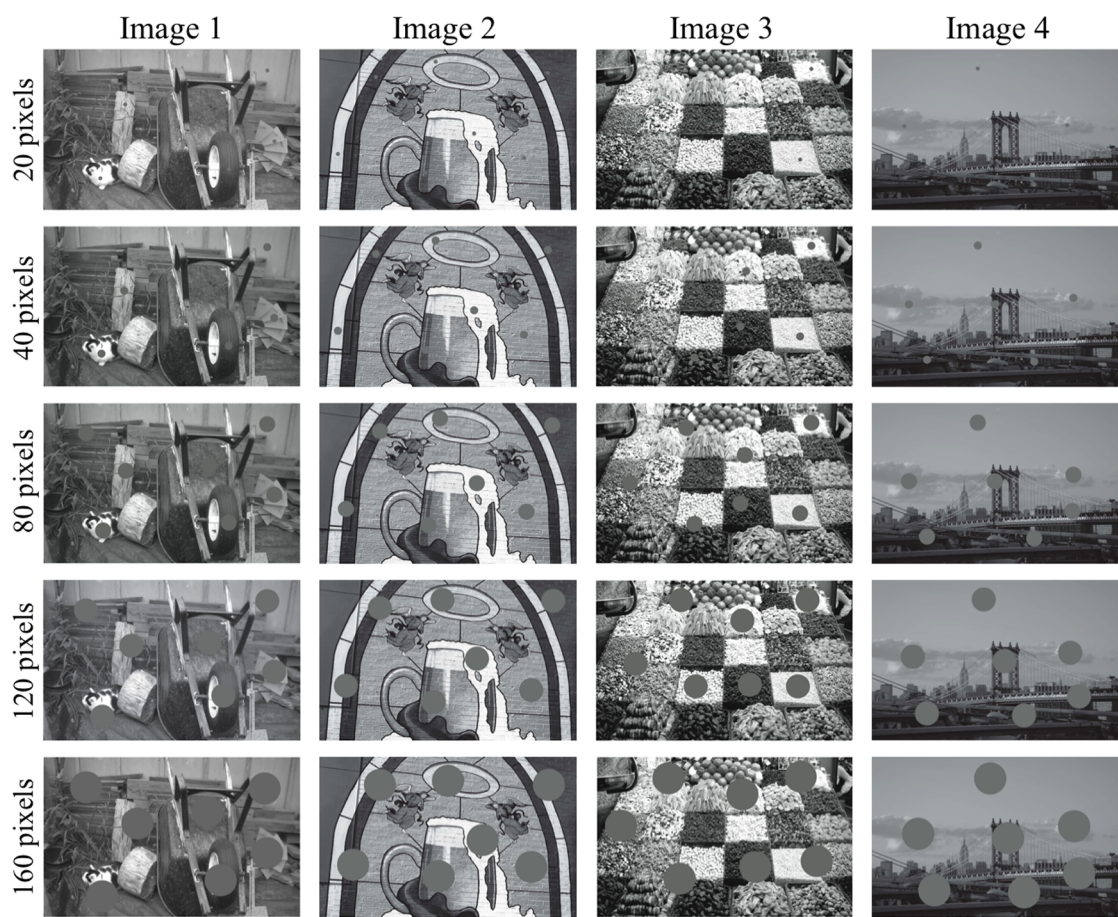


Figure 2 | The four grayscale images used in the study. Each image contains seven test disks (pixel value 127). Each row shows all four images with a particular size of test disk. The test disks range in diameter from 20 pixels (top row) to 160 pixels (bottom row). Observers viewed the full set of images at four distances from the computer screen: 50, 100, 200, and 300 cm. The task was to rank the test disks in each image from darkest to brightest. All images were taken by Erica L. Dixon.

Consistent with these observations, a one-way ANOVA that tested the equality of observer rankings for each disk across viewing distance showed few significant differences; only two out of twenty-eight conditions in Image 1 had $p < .05$, disk B at 80 pixels and disk B at 160 pixels (see table 1 for ANOVA values of significant differences across all images).

Correlation between observer rankings and physical values in filtered image. We analyzed the images using a high-pass filter, the goal of which is to calculate the correlation between observer rankings (average observer rankings of test disks on unfiltered image) and the measured pixel values of the test disks after the original image has been filtered.

The only parameter in equation (1) is the size of the averaging kernel (H). Adjusting the size of the kernel adjusts the amount of spatial frequency content available in the final image. A small averaging kernel creates blur only over a small region in the convolved image (i.e., Original Image* H), therefore, only sharp edges will remain in the final image when the blurred image is subtracted from the original image. A larger averaging kernel creates a larger blur area, and as result, a wider range of low-spatial frequency content will remain when the blurred image is subtracted from the original image.

In our analysis we parametrically changed the size of the kernel (kernel sizes ranged from 5 pixels in diameter up to 1000 pixels in diameter), measured the pixel values of the tests disks in the filtered image, and then calculated the correlation between the ranking of the pixels values and the observer brightness rankings. Figure 4 shows an example of the correlations plotted as a function of filter size; there

were 80 such plots, one for each disk size at each of four distances for all four images. To estimate the filter size that produced the peak correlation, we fit a Gaussian function to the data in each of the correlation vs. filter size plots; the fit is shown as a red line in Figure 4.

The peak correlations for each image and each pixel size, and the corresponding filter kernel size, are shown in Table 2 (for clarity we averaged across distance for each pixel size). The maximum correlation values were above 0.85 for disk diameters of 80 pixels and below; the correlations decline for most larger disk diameter conditions, perhaps because perceived differences between the disks are less apparent on average for larger disks (i.e. the illusory effect is not perceived). The clear peak in correlation value versus filter size was shown for Images 1, 2 and 3. For Image 4, the correlation values did not decline but remained high for all filter values greater than the size of the disk—this might be expected for Image 4, since the image has fewer brightness transitions (i.e., the image is mostly sky or bridge).

How does the best filter size change as a function of distance and disk size. The goal of the analysis is to estimate the effects of changing the disk size and viewing distance on the size of the optimal filter size. If the filter depends on relative object-size in the image rather than retinal size then the optimal filter size should increase as the disk size increases and remain constant as a function of viewing distance. Figure 5 plots the best filter size as a function of test disk size; each panel shows the results for a different image; and each symbol represents a different viewing distance. In order to estimate the best filter size, we fit Gaussian functions to individual observer correlation plots (i.e., plots that are similar to Figure 4, but for individual

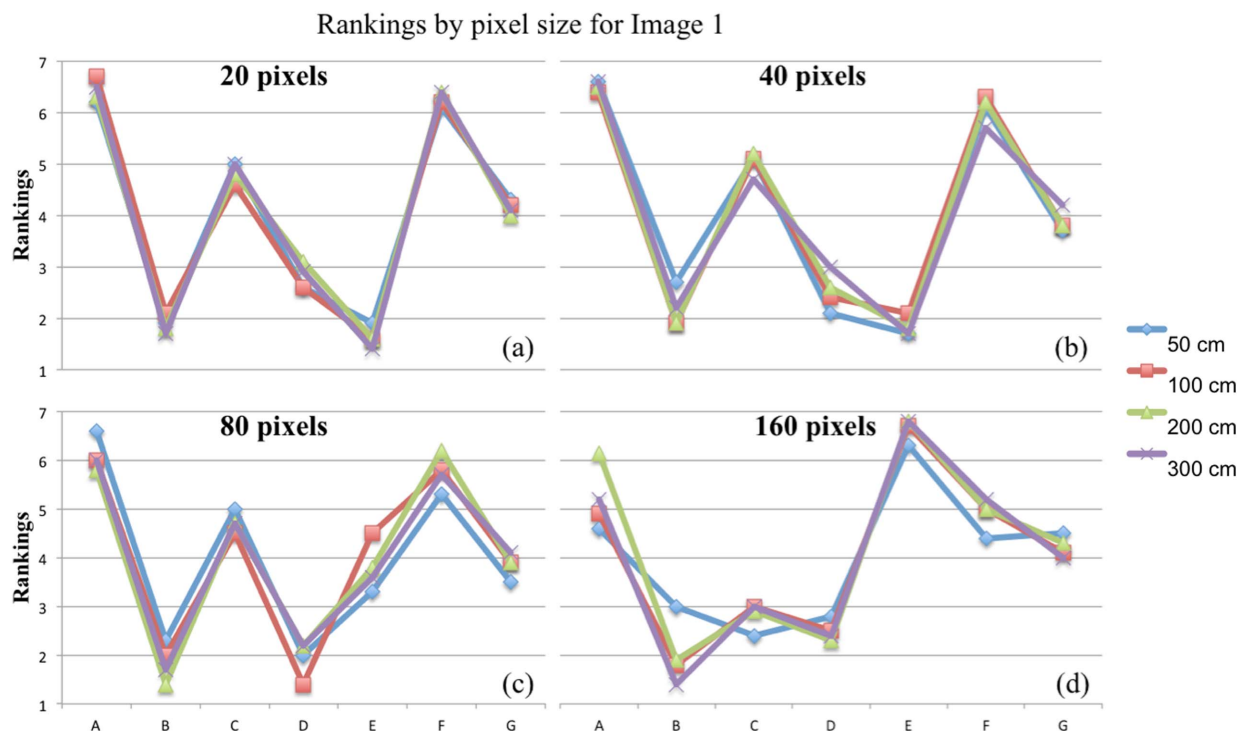


Figure 3 | Average observer rankings for Image 1. Each panel shows results for different sizes of test disk diameter (panel a = 20 px, b = 40 px, c = 80 px, d = 160 px); each line represents observer viewing distance (blue = 50 cm, red = 100 cm, green = 200 cm, purple = 300 cm). The y-value shows the average ranking for each test disk across all observers.

observers) giving us a set of ten peak filter size values for each condition. We then used a bootstrapping procedure with 1000 iterations in Matlab to estimate the standard error of the best fitting filter sizes (Figure 4).

There are two main observations from the plot in Figure 5: 1) As disk size increases, so does the optimal filter size; for all three curves the optimal filter follows a line approximately equal to $0.35 \ln(x) + 0.8$. 2) The optimal filter size is not affected by changing viewing distance since in each panel of Figure 5, data from all viewing distances cluster together at the peak filter size. We excluded Image 4 from the figure because, as stated above, a Gaussian curve could not be fit to the correlation vs. filter plots because all filter sizes above the size of the disk diameters produced very strong correlations (i.e., for this figure most filter sizes worked). We have shown that ranking changes as a function of object size in the image and that, for any object size, the ranking remains a relatively constant function of viewing distance. The results suggest that brightness effects are viewing distance invariant.

Study 2: Extension of model to Knill and Kersten Illusion (1991).

How well can the simple filter account for a well-known illusion that seems to suggest brightness estimates require inferences about the illumination? In the Knill and Kersten¹⁸ illusion, two identical shaded gradients appear dramatically different when viewed as the front

surface of two cubes, but appear similar to each other when viewed as the front surface of cylinders (Figure 6a). The proposed theoretical explanation is that when the gradients are cylinders, the visual system infers that the dark regions are in shadows created when the central part of the cylinders block the illumination; when the gradients are interpreted as a flat face of the cubes, no such inference is possible. We tested whether the Shapiro-Lu model could account for the perceptual disparities between the cubes and cylinders.

In order to determine if the filter could produce an image that corresponds to the perceived appearance of the illusion, we filtered the image with a series of convolution kernel diameters ranging in size from 60 pixels to 240 pixels. Figure 6(b–e) shows the images after filtering at consecutively larger filter sizes, up to 240 pixels—the diameter of the paired shaded gradients; panels g–j show the measured pixel value compared to the unfiltered image. When the filter size is scaled to the shaded gradients, i.e. 240 pixels, (Figure 1d and 1i), the object-level filtered image physically mirrors the brightness illusion demonstrated in the original.

The Shapiro-Lu high-pass filter model can account for the brightness differences between the cylinders and the cubes in the Knill and Kersten illusion, providing an alternate explanation for the illusion; a high-pass filter creates brightness differences and does not require the visual system to make unconscious inferences about illumination. In order for the filter to make the appropriate predictions, however, the size of the blur kernel must be adjusted to the approximate size of the gradient fields. This result once again suggests that the cut-off frequency of the filter corresponds to the size of the attended object.

Discussion

Here we examined how the Shapiro and Lu⁶ filter model accounts for observer rankings of multiple sizes of test disks placed within natural images and viewed at a range of distances. As a general rule, the relative brightness of the test patches did not change as a function of viewing distance. We have replicated the Shapiro and Lu finding

Table 1 | ANOVA

Across Distance						
	Image	Pixel Size	Disk	df	F	Sig
	1	80	B	3	2.90	0.05
	1	160	B	3	4.58	0.01
	2	40	E	3	3.05	0.04
	3	40	E	3	3.05	0.04
	3	80	E	3	4.81	0.01

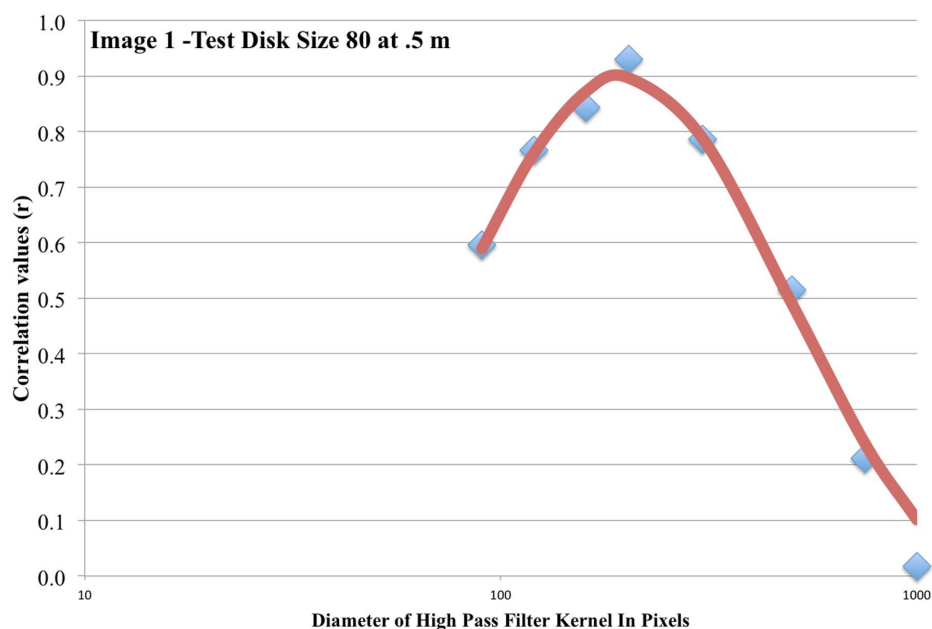


Figure 4 | Correlation between observer rankings and test disk pixel values. We calculated the correlations between observer rankings and the physical values of test disks in the filtered image. The figure shows correlations as a function of filter size in equation (1) for Image 1, test disk size 80, at viewing distance of 50 cm. Such calculations were performed for each image at each test disk pixel size and viewing distance (80 conditions total). We fit a Gaussian function to each plot, seen as a red line. The peak filter size for each condition was taken as the mean of the fitted function.

that such brightness rankings can be accounted for by simply removing low spatial frequency content from the image, indicating that even though all the test disks have the same pixel value, the test disks are actually physically different from each other after high-pass filtering at some spatial scale. For Images 1, 2, and 3, only a narrow range of filters could account for the observer rankings: intermediate-sized filters produced larger correlations to observer rankings, while large and small filters produced zero correlations. The results indicate that the visual system removes a greater range of low spatial frequency content when the test disks are small than when the test disks are large. A static filter that responds to a fixed retinal size or retinal frequency would not be able to account for these results, instead, the results require a filter tuned to the absolute size of the attended object.

We used test patches placed in natural images based on Gilchrist and Radonjic⁴. An advantage of using natural images can be seen in Image 4. In this image, unlike the other three, nearly all filter sizes larger than the size of the disks “worked”; that is, a wide range of filter sizes produced strong correlations with observer rankings. Image 4 is therefore consistent with tuned filter responses but also with models in which the filter remains fixed. Image 4 was different from the other three images in that it contained fewer brightness transitions. The results suggest that filter tuning might be psychophysically detectable only in relatively complex scenes, and that spatially complex scenes may be required to test differences between brightness models. For instance, others have shown that brightness illusions are relatively

independent of viewing distance¹⁵, and brightness invariance with distance can be accounted by several filtering-based models^{16,19,20}. While multi-scale- and static-filter approaches can account for many brightness illusions^{9,16,17,21}, the difference between Images 1–3 and Image 4 suggests that it may be worth examining these models with spatially complex images as well, since such images may test how well the models perform in the presence of a wider variety of spatial information.

Any model that adaptively removes low spatial frequency content will be able to account for most brightness illusions, even when the images are viewed at a variety of distances. For instance, in multi-channel models, the channel with maximum response is accentuated relative to the other channels by some form of divisive normalization and thereby lowering the response from the low spatial frequency channel (in most conditions)^{22,23}. The question we ask is whether the weighting of spatial channels allow us to perceive brightness illusions, or does brightness follow from a weighting function that is part of a broader, more functionally important, role? Multichannel processes have to serve several different functions of the visual system (multiple motion systems, color, texture segregation, object form, object and face identity, etc.), and each of these undoubtedly requires weighting of the spatial channels that correspond to the tasks that they are performing. There are several other processes that could change the spatial weighting function. For instance, much of the filtering does not have to do with brightness or gain control, per se, but rather with compensation for eye movements to prevent motion blur. Eye movements cause suppression of low spatial frequencies carried by the magnocellular pathway²⁴, and ganglion cells’ spatial responses shift towards higher spatial frequencies^{14,25}. The effects of eye movements on spatial frequency seem to follow the reduction of low spatial frequency response, and could therefore create a weighting function that may be similar to those needed to account for relative brightness perception.

Another possibility – one that we favor – is that one of the purposes of multiple spatial channels is to create an object-level representation of the world. If the disks in our displays can be considered visual objects, then there is an invariant relationship between the disks and the background. Invariance of this sort has been suggested

Table 2 | Peak Correlations and Filter Size

Pixel size of disks	Image			
	Image 1	Image 2	Image 3	Image 4
20	0.97 (20)	0.96 (20)	0.9 (40)	0.98 (40)
40	0.94 (160)	0.95 (80)	0.89 (80)	0.97 (80)
80	0.85 (240)	0.94 (240)	0.85 (120)	0.96 (240)
120	0.63 (240)	0.94 (240)	0.85 (120)	0.97 (240)
160	0.17 (240)	0.82 (500)	0.61 (160)	0.96 (240)

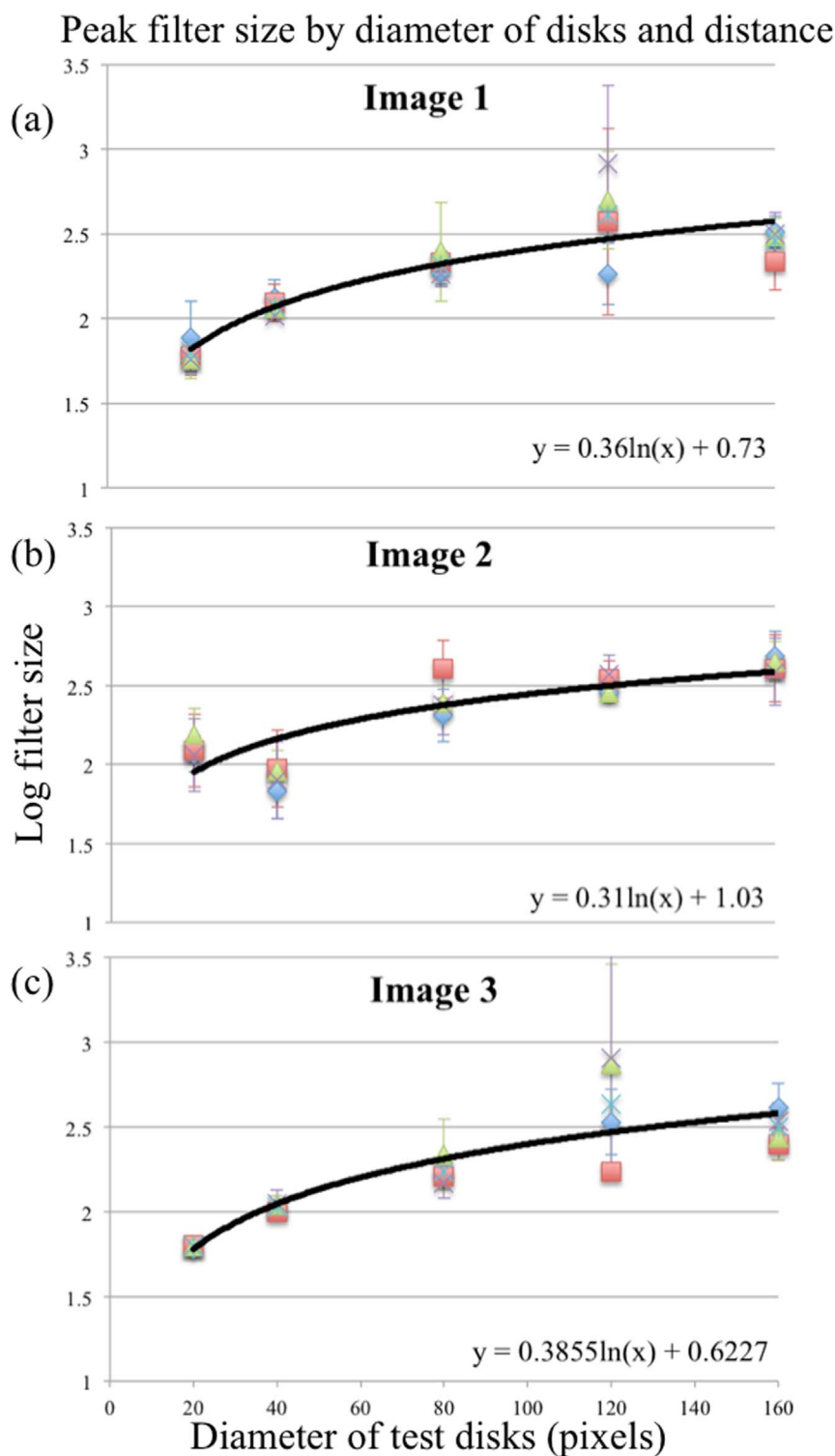


Figure 5 | Peak filter size as a function of test disk size. We calculated the peak filter size (see Figure 4) for each condition. Here we plot the change in filter size as function of test disk diameter. If the filter depends on object size rather than retinal size then the optimal filter size should increase as the disk size increases and remain constant as a function of viewing distance. Panels (a–c) show the log filter size versus the diameter of the test disk for Images 1–3; Image 4 not shown as a Gaussian could not be fit to the correlation plots as seen in Figure 4 (see Discussion). Error bars are $\log(y) \pm dy/y$ of the bootstrapped estimate of variance. The solid lines are the best-fit regression lines to the data; x is disk diameter.

in some psychophysical studies that have shown that the crucial variable for object detection is not retinal spatial frequency, but object spatial frequency relative to the image spectrum¹³. Furthermore, much of visual cognitive neuroscience literature concerns separate processes for object perception^{26,27,12}. Roe et al.²⁸ recently

proposed a theory concerning the functional purpose of V4—a cortical area in the early stages of the ventral visual pathway, suggesting that V4 combines brightness and other cues to enhance “figureness” by differential neuronal response to objects and their surrounds. Such an approach is consistent with idea that that some form of



Knill and Kersten Illusion

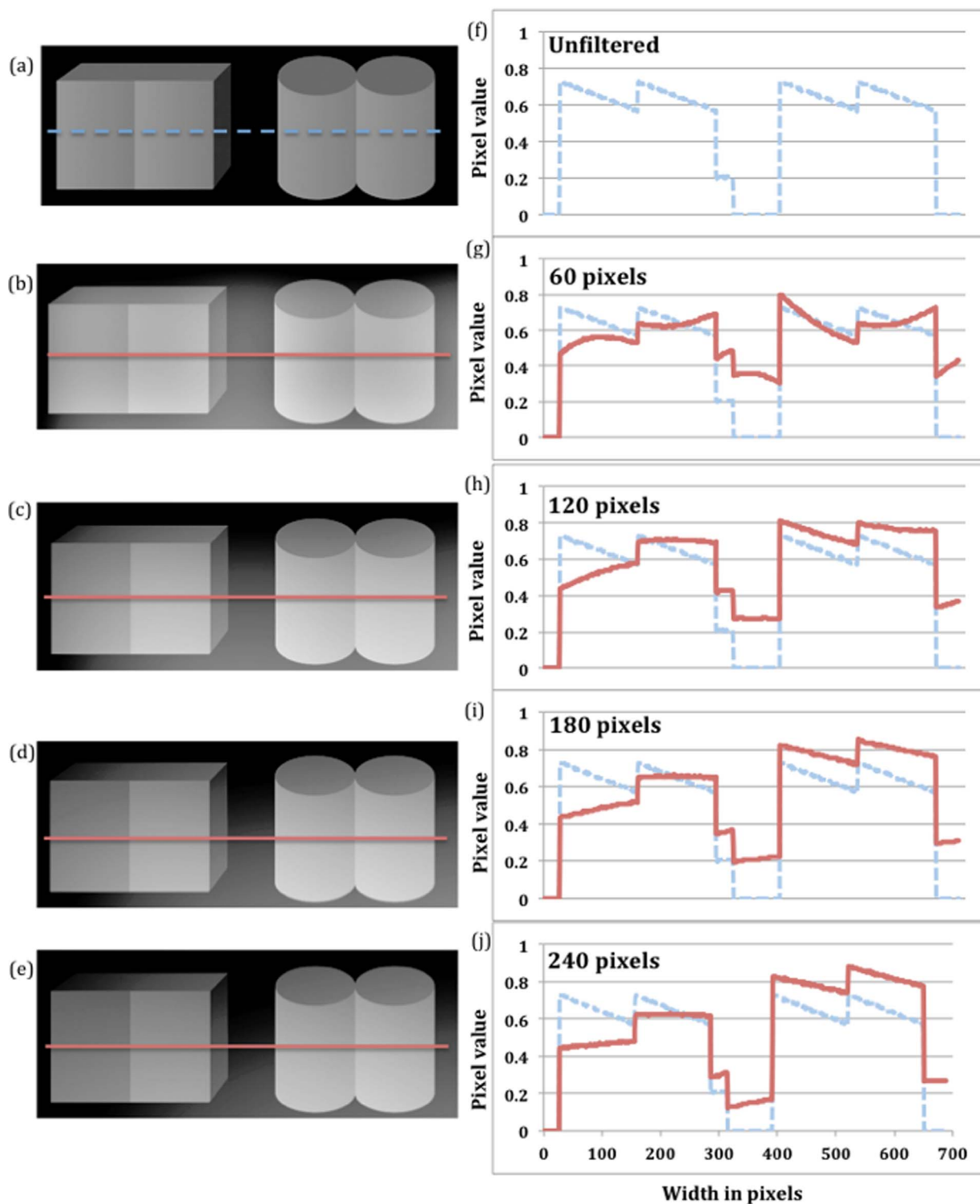


Figure 6 | Filter model applied to Knill Kersten (1991) illusion (a) Knill and Kersten illusion: test gradients are identical luminance levels; the gradients for the cubes appear different from each other, gradients for the cylinders appear similar. Panels (b–e) show high-pass filtered versions of the illusion with increasing size of the kernel diameter (in filtered image, pixel size of each gradient was 120 pixels). Panels (f–j) show pixel level profiles for the images (a value of 1 is the highest pixel level, i.e., 255): the blue dashed line indicates the level for the unfiltered image (shown by itself in panel f); the red solid line indicates the pixel level for the image in the corresponding row. Filter sizes 120 to 240 show profiles corresponding to the perception of the brightness in the unfiltered image.



visual representations are encoded as object files¹⁰. Object files can only be possible if the object representation has removed the information that conveys illumination and shadows within a scene and other information that is irrelevant to an object's content. An object representation will, almost by default, create a reduction of low spatial frequency content similar to those reported here (and implicitly by other spatial-frequency-type models of brightness perception). Indeed, our results suggest that simply by separating a figure from the ground, the visual system may be triggering events that select for higher spatial frequency content (or that do not respond to low spatial frequency content) and therefore may be producing what is commonly thought of as a brightness illusion.

Adjusting filters to the size of the object could possibly account for brightness changes that occur without changes in the visual image. For instance, it has been shown that spatial organization can affect the extent of brightness illusions²⁹ (another compelling illusion in this vein was recently presented by Hong and Kang³⁰). Tse³¹ showed that simply shifting attention from one disk to another while maintaining constant fixation could change the brightness levels of three identical overlapping disks. Both of these results would be hard to account for strictly by bottom-up processes with contrast normalization; however, an object-level filter approach predicts that changes in perceptual organization would lead to changes in brightness since a larger or smaller grouped object would lead to shifts in the filter cut-off. Also, it would not be surprising if attended objects create a finer perceptual representation than unattended objects; to produce a finer representation, the visual system would have to exclude more low spatial frequency responses, which would lead to a change in brightness perception. This type of process is consistent with other findings of the effect of attention on spatial frequency responses³².

Indeed, the object-level approach provides a response to a puzzle raised by Paul Whittle: the observation that “colour is always perceived relative to its background [in brightness illusions] is contradicted by the everyday observation that if you move an object against a variegated background, it is often hard to see any changes in its colour at all³³.” If objects are the fundamental level of interest, then such problems should be easily accounted for, since the size of the filters adjusts to the objects in the scene. This is particularly true if one considers the role of object layering in brightness perception; Anderson and Winawer^{34,35} have been strong advocates for the role of scission in perceptual interpretations. In many respects, the argument in favor of filtering by object size is consistent with the central tenets of the argument for scission, since scission is essential for object formation. The major advantage of an object-based representation is that scission layers by themselves do not necessarily indicate the size of the filter.

Much of the literature related to unconscious inference theories assumes that the visual system attempts to “discount the illuminant” so as to estimate the reflectance of the surface, and assume, implicitly or explicitly, that features of objects are essential for understanding brightness/lightness³. As we have shown, in the case of the Knill and Kersten illusion, an object-based filter that removes low spatial frequency content may be thought of as serving to discount the illuminant, as well; indeed, at a practical level, the Adobe Photoshop high-pass filter is frequently used to reduce the effects of shadows while maintaining image detail and to remove shading patterns introduced into textures. The filter in equation (1) would serve the purpose of reducing the effects of illumination changes, therefore allowing the visual system to make a better estimate of surface reflectance^{36,37}. The advantage of a filter technique is that the visual system would be making these inferences based upon the information present in the image and would therefore not require knowledge about the illumination in the scene.

Our approach does not eliminate the need for linking rules such as those found in anchoring theory³⁸, which creates an explicit rule for assigning lightness to levels of luminance value. We do note, how-

ever, that any such linking rule is likely to be based on a high-pass filtered version of the image, not on an analysis of the pixel values or individual points in the image. In addition, we speculate that the size of the relevant frameworks within a scene would influence the size of the filter. Anchoring theory makes clear predictions for how an object should appear depending on its specific perceptual framework. When an object changes its framework (either through an act of the observer, through motion³⁹, or through changing its depth plane^{40,41}), we would also expect the size of the filter to change, thus producing a change in relative appearance. We have not yet tested whether the size of the filter would produce relative value changes consistent with those expected from anchoring.

Lastly, one recent approach to brightness from Dale Purves²⁵ laboratory is that our perceptual world is empirically based on our past experience with surfaces and illuminants. A major tenet of this theory is that perceptions stem from the process of connecting retinal images with successful and valuable behaviors. While it is certainly likely that experience influences brightness, a high-pass filter can account for relative brightness changes in most of the very impressive brightness demonstrations included in Purves²⁵ research. Our results suggest that rather than learning a complex range of possible illuminations and surface reflectances, the visual system would learn to select the appropriate channels for producing an object. Once the object is perceptually defined, much of relative brightness perception is a given, and many illumination problems become easier to handle.

In conclusion, we have replicated our previous findings that a simple filter that removes low spatial frequency content can account for relative brightness rankings of test spots in natural scenes once the filter is adjusted for the size of the object. As stated in Shapiro and Lu (2011), the reason for this is that in most brightness illusions, test patches with identical pixel values are actually physically different from each other when considered at the appropriate spatial scale. Any theoretical approach that, in effect, removes low spatial frequency content from the image will therefore in principle account for simultaneous contrast phenomena. Furthermore, we note that in the natural environment, lightness and brightness are usually attached to objects; a representation of a visual object does not need to include spatial frequency content that is lower than the size of the object. Object representations, therefore, act as the appropriately sized spatial filter to produce the effects demonstrated in this paper. Object identification occurs rapidly and is probably the end result of many processes dedicated to extracting spatially invariant objects from the visual image¹². So, while spatial filtering occurs at many different stages of processing, it is likely that representations of visual objects are constructed of information that is subsequently required for the production of simultaneous contrast phenomena.

Methods

Observers. Ten undergraduate and graduate students at American University with normal or corrected-to-normal vision participated.

Materials. To measure perceived test disk brightness in images of natural scenes, we presented observers with a set of twenty images comprised of four grayscale photographs (1856 × 1160 pixels), reproduced five times. Each image contained a single size of seven identical mid-luminance-level test disks (the diameters were 20, 40, 80, 120, or 160 pixels); the complete set of images is shown in Figure 2. Each image had a midscale gray border extending to the edges of the computer screen to ensure contrast between the edges of each picture remained constant and neutral.

Images were presented on a 27" iMac LCD screen set to a linear gamma level of 1.0. A uniform 127 value gray 8 × 4 grid matched to the size of the presentation images was measured using a photometer at 32 locations; luminance values varied from 0–15%. We created a filter in Matlab to increase or decrease the value of each section of the image to ensure that while the pixel value was not identical, the luminance values were much less varied; variance for the filtered gray grid ranged from 0–2% from the center average. The filter was used to adjust each photo. Four randomized presentation series were created to ensure that participants viewed the twenty images in a novel order at each of the four distances. Additionally, the order of the four presentations was arranged to create four distinct viewing orders that varied among participants.



Procedure. Observers viewed the images at four distances from the computer screen: 50, 100, 200, and 300 cm. The task was to rank the disks from darkest to brightest—so, for instance, in Figure 1a, most observers would rank the disk labeled E as 1 since they perceive it as darkest, and the disk labeled A as 7 since they perceive it as brightest. Each participant viewed a practice presentation series to ensure understanding of the viewing and ranking process; the number of images viewed by each participant varied based on comfort with the response system. Rankings were recorded on paper containing a schematic replicating the arrangement of the test disks on the images. After all disks were ranked, the experimenter advanced to the next image until all twenty images had been completed; no time limit was placed on responses. After the completion of each series, the participant moved to the next distance to rank the same images in a novel order.

- Gilchrist, A. L. *Seeing black and white*. (Oxford University Press, 2006).
- Kingdom, F. in *Levels Percept.* (Harris, L. & Jenkin, M.) 19–42 (Springer, 2003).
- Adelson, E. H. Lightness perception and lightness illusions. *New Cogn. Neurosci.* 339–351 (MIT Press, 2000).
- Gilchrist, A. L. & Radonjić, A. Functional frameworks of illumination revealed by probe disk technique. *10*, 1–12 (2010).
- Purves, D., Williams, S. M., Nundy, S. & Lotto, R. B. Perceiving the intensity of light. *Psychol. Rev.* **111**, 142–58 (2004).
- Shapiro, A. & Lu, Z.-L. Relative brightness in natural images can be accounted for by removing blurry content. *Psychol. Sci.* **22**, 1452–9 (2011).
- Zaidi, Q., Yoshimi, B., Flanagan, N. & Canova, A. Lateral interactions within color mechanisms in simultaneous induced contrast. *Vision Res.* **32**, 1695–1707 (1992).
- Shapley, R. & Reid, R. C. Contrast and assimilation in the perception of brightness. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 5983–5986 (1985).
- Perna, A. & Morrone, M. C. The lowest spatial frequency channel determines brightness perception. *Vision Res.* **47**, 1282–91 (2007).
- Kahneman, D., Treisman, A. & Gibbs, B. J. The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* **24**, 175–219 (1992).
- Kersten, D., Mamassian, P. & Yuille, A. Object perception as Bayesian inference. *Annu. Rev. Psychol.* **55**, 271–304 (2004).
- Isik, L., Meyers, E. M., Leibo, J. Z. & Poggio, T. a. The dynamics of invariant object recognition in the human visual system. *J. Neurophysiol.* (2013) doi:10.1152/jn.00394.2013.
- Parish, D. H. & Sperling, G. Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Res.* **31**, 1399–415 (1991).
- Rucci, M., Iovin, R., Poletti, M. & Santini, F. Miniature eye movements enhance fine spatial detail. *Nature* **447**, 851–854 (2007).
- Foley, J. M. & McCourt, M. E. Visual grating induction. *J. Opt. Soc. Am. A Opt. image Sci.* **2**, 1220–1230 (1985).
- Blakeslee, B. & McCourt, M. E. A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. *Vision Res.* **39**, 4361–77 (1999).
- Blakeslee, B., Pasioka, W. & McCourt, M. E. Oriented multiscale spatial filtering and contrast normalization: a parsimonious model of brightness induction in a continuum of stimuli including White, Howe and simultaneous brightness contrast. *Vision Res.* **45**, 607–15 (2005).
- Knill, D. C. & Kersten, D. Apparent surface curvature affects lightness perception. *Nature* **351**, 228–30 (1991).
- Olzak, L. A., Laurinen, P. I. & Peromaa, T. L. Early Cortical Influences in Object Segregation and the Perception of Surface Lightness. *Psychol. Sci.* **8**, 386–390 (1997).
- Robinson, A. E., Hammon, P. S. & de Sa, V. R. Explaining brightness illusions using spatial filtering and local response normalization. *Vision Res.* **47**, 1631–1644 (2007).
- Dakin, S. C. & Bex, P. J. Natural image statistics mediate brightness “filling in”. *Proc. Biol. Sci.* **270**, 2341–8 (2003).
- Blakeslee, B. & McCourt, M. E. When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision Res.* **60**, 40–50 (2012).
- Robinson, A. E. & De Sa, V. R. Brief presentations reveal the temporal dynamics of brightness induction and White’s illusion. *Vision Res.* **48**, 2370–2381 (2008).
- Burr, D. C. Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature* **371**, 511–513 (1994).
- Kuang, X., Poletti, M., Victor, J. D. & Rucci, M. Temporal Encoding of Spatial Information during Active Visual Fixation. *Curr. Biol.* **22**, 510–514 (2012).
- Valyear, K. F., Culham, J. C., Sharif, N., Westwood, D. & Goodale, M. A. A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal visual streams: a human fMRI study. *Neuropsychologia* **44**, 218–228 (Elsevier, 2006).
- Kanwisher, N., Chun, M. M., McDermott, J. & Ledden, P. J. Functional imaging of human visual recognition. *Brain Res.* **5**, 55–67 (1996).
- Roe, A. W. *et al.* Toward a Unified Theory of Visual Area V4. *Neuron* **74**, 12–29 (2012).
- Xian, S. X. & Shevell, S. K. Changes in color appearance caused by perceptual grouping. *Vis. Neurosci.* **21**, 383–388 (2004).
- Hong, S. W. & Kang, M.-S. Perceptual consequence of normalization revealed by a novel brightness illusion. [abstract]. *J. Vis.* **12**, article 1219 (2012).
- Tse, P. U. Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Res.* **45**, 1095–1098 (2005).
- Carrasco, M., Loula, F. & Ho, Y.-X. How attention enhances spatial resolution: evidence from selective adaptation to spatial frequency. *Percept. Psychophys.* **68**, 1004–1012 (2006).
- Whittle, P. Contrast colours in *Colour Percept.* (Mausfeld, R. & Heyer, D.) (Oxford University Press, 2003).
- Anderson, B. & Winawer, J. Image segmentation and lightness perception. *Nature* **434**, 79–84 (2005).
- Anderson, B. L. & Winawer, J. Layered image representations and the computation of surface lightness. *J. Vis.* **8**, 18.1–22 (2008).
- Zhang, M. How to Use Photoshop’s High Pass Filter to Soften Skin While Retaining Texture. (2011).
- Hajba, P. The Power of the High Pass Filter. (2001).
- Gilchrist, A. *et al.* An anchoring theory of lightness perception. *Psychol. Rev.* **106**, 795 (1999).
- Werner, A. Color constancy improves, when an object moves: high-level motion influences color perception. *J. Vis.* **7**, 19.1–14 (2007).
- Gilchrist, A. L. Perceived Lightness Depends on Perceived Spatial Arrangement. *Science*. **195**, 185–187 (1977).
- Gilchrist, A. L. When does perceived lightness depend on perceived spatial arrangement? *Percept. Psychophys.* **28**, 527–538 (1980).

Acknowledgments

Supported by NIH grant R15EY021008 to A.G.S. The authors thank Sherri Geller for her editorial assistance.

Author contributions

E.L.D. and A.G.S. designed and conducted the experiment, collected the data, analyzed the data, and wrote the manuscript. Z.-L.L. edited the manuscript and gave conceptual advice.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Dixon, E., Shapiro, A. & Lu, Z.-L. Scale-invariance in brightness illusions implicates object-level visual processing. *Sci. Rep.* **4**, 3900; DOI:10.1038/srep03900 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>