A cue-free method to probe human shape from shading

Mazzilli, Giacomo; Schofield, Andrew

DOI:
10.1068/p7517

Perception, 2013, volume 42, pages 932–940

doi:10.1068/p7517

A cue-free method to probe human lighting biases

Giacomo Mazzilli, Andrew J Schofield
School of Psychology, University of Birmingham, Edgbaston Campus, Birmingham B15 2TT, UK;
e-mail: gxm947@bham.ac.uk
Received 1 May 2013, in revised form 23 September 2013

Abstract. People readily perceive patterns of shading as 3-D shapes. Owing to the generalised bas-relief ambiguity when extracting shape from shading, people must simultaneously estimate the shape of the surface and the nature of the light source. In many cases cues in the image will be insufficient to resolve all of the ambiguities present, and in such cases the human visual system may employ one of a number of prior assumptions based on ecology and experience. One such assumption is the lighting-from-above prior. Here, in the absence of extrinsic cues to lighting direction, ambiguous shading patterns are interpreted as if lit by a light source that is above the observer's head. Studies of this prior typically use ambiguous stimuli and observe perceptual biases. A degree of cueing is inherent to such methods. Participants see the shaded stimuli repeatedly and are asked to make shape judgments about them regardless of whether or not they actually perceive any 3-D shape. We wanted to access people's lighting prior more directly by establishing the template they would employ to detect a shaded object in the absence of *any* visual cue to object shape. To this end, we adopted a classification image approach.

Keywords: 3-D shape perception, visual perception, classification images, shape from shading, visual psychophysics

1 Introduction

When presented with ambiguous stimuli, the human visual system must decide on a percept by making assumptions about the scene which may not be correct. This is the case in figure 1, where a concave dip lit from below is often perceived as a convex bump lit from above.

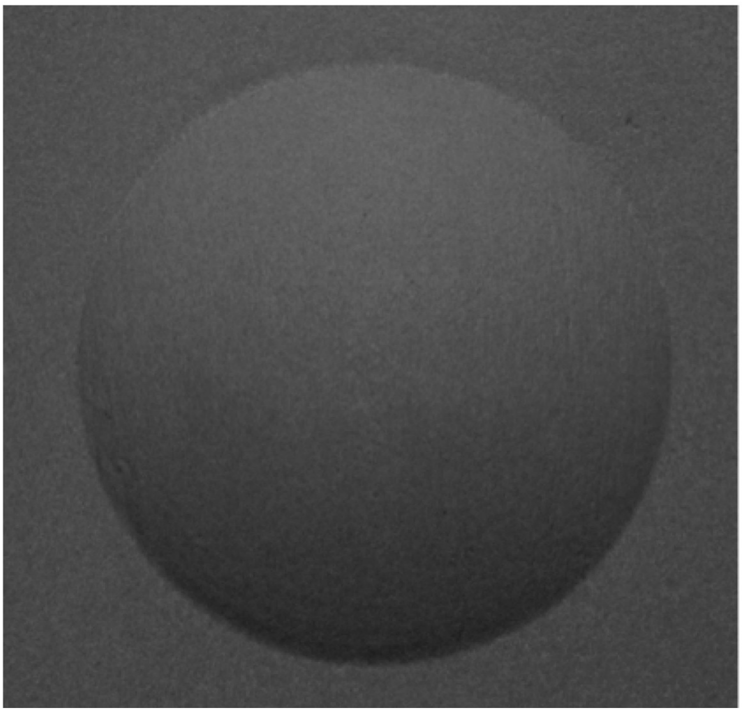

Figure 1. A photograph of a physical concavity lit from below which appears as a convexity lit from above.

In the absence of extrinsic cues to lighting direction, ambiguous shading patterns are often interpreted as if lit by a light source that is above the observer's head (Adams et al 2004; Brewster 1826; Gerardin et al 2010; Mamassian and Goutcher 2001; Ramachandran 1988; Rittenhouse 1786; Sun and Perona 1998; Todd 2004). However, the lighting-from-above prior is not the only lighting assumption that humans adopt (Langer and Bülthoff 2000; Schofield et al 2011; Tyler 1997), and lighting assumptions can be overridden by assumptions about object shape (Liu and Todd 2004), and by lighting cues in the image itself (Morgenstern et al 2011). In many studies showing robust lighting-from-above priors the stimulus is clearly shaded by a strong directional light source but in an ambiguous manner—such that the observer is forced to choose between two interpretations, one of which happens to be consistent with lighting from above while the other is not. In such cases, where the assumption of a directional light source is encouraged over a diffuse interpretation, lighting from above 'wins', but it may not be so robust in other circumstances. Recent studies (Liu and Todd 2004; Morgenstern et al 2011) have demonstrated how easily the lighting-from-above prior can be overridden. We test the robustness of the lighting-from-above prior using stimuli that have no shading structure and no lighting cues. Thus, we directly probe the observers' internal templates or assumptions for interpreting shaded objects in the absence of any interference or bias from stimulus features.

We address this question using classification images (CIs) as a tool to probe the lighting prior. This technique [see Murray (2011) for a review] has been used in psychophysics in order to elucidate the nature of templates or filters within human vision by correlating observers' decisions with apparently uninformative noise features over a large number of trials. It has been used to investigate problems such as Vernier acuity (Ahumada 1996), illusory contours (Gold et al 2000), and letter discrimination (Watson 1998). The classical (CI) approach has been extended to include stimuli composed of noise-only images (Gosselin and Schyns 2003). In this study observers were instructed to detect the presence of a target in white noise; no signal was ever added to the white noise patch, but observers were led to believe that the stimuli included a target signal. As with the classical method, the resulting CI represents the template used by the participant to perform the task. If the observer does not apply a systematic approach, the resulting template should have the same statistical properties as the noise. A systematic approach would be indicated by structure in the template, and this structure reveals both the observer's impression of what the target should look like given the instructions, and their strategy for detecting it. The use of noise-only stimuli is crucial to our experiment, as it is impossible for such stimuli to bias perception. For each observer the resulting template would not only reveal what the target should look like—in this case what a bump or a cylinder or disk look like—but would also have embedded information about light source directions. For instance, we use $1/f$ noise as it appears to present more plausible image forms and therefore to make it easier for observers to believe that a target was actually embedded in the noise (even though there was no target). We do not think that this type of noise engages the visual mechanisms that recover shape from shading. Instead, it facilitates the task, allowing us to directly measure people's templates for interpreting shaded surfaces.

Observer's templates were derived by accumulating the noise samples leading to positive responses and subtracting the accumulation of the unselected noise samples. Templates for shaded bumps should contain a highlight indicating the part of the imagined surface that is directed towards the imagined light source. If observers applied the lighting-from-above prior, we should expect such highlights to lie above the central row of their templates. To test the dependence of the template on task demands, we tested three control conditions: large bumps (diameter 4 deg), cylinders (2.5 deg × 4 deg), and white disks (2.5 deg). We expected to find larger offsets for highlights in the large bump condition; horizontally elongated templates for the cylinder condition; and flatter templates with less distinct highlights and no offset in the white disks condition: flat white disks do not imply any shading highlight.

Templates were typified by approximately Gaussian blobs, which varied systematically with the instructions given. To reveal observers' lighting bias, we fitted a Gaussian to the central column of each template and took the position of its peak as the location of the illumination highlight that observers where aiming to detect in the stimulus. We judged the effective lighting prior by measuring the distance of the peak from the middle of the template.

2 Method

2.1 *Participants*

Seven naive observers participated in the small bump condition; four of these also took part in the large bump condition; two observers participated in the white disk condition and one in the cylinder condition. All observers had normal or corrected-to-normal vision. All participants were students at University of Birmingham, except TY and HB who were employees; all except the employees were paid £6 per hour. They all gave their consent to take part in the study.

2.2 *Stimuli*

All stimuli were noise samples with an approximate $1/f$ amplitude spectrum in order to simulate the spectral content of natural scenes without introducing systematic structural features; they were created and displayed in monochrome (grey) using Visual C++ 6.0 (Microsoft). Stimuli were presented on a Sony 520GDMF monitor using a VSG2/5 graphics card (CRS Ltd, Rochester, UK) in a dark room in order to eliminate any cue to lighting direction. The image size was 10 × 10 deg at the viewing distance of 114 cm (resolution = 512 × 512 pixels). The monitor's gamma nonlinearity was estimated using a CRS-ColourCal Photometer and corrected using lookup tables in the VSG.

2.3 *Procedure*

We used a two-interval forced-choice (2IFC) paradigm in which observers were told that a target would be added to the noise in one of two intervals chosen at random and were asked to identify the interval containing the target. However, no targets were ever presented. Rather, the two presentation intervals contained noise-only images and the participant indicated which one 'looked most like' an imagined target. Each trial consisted of 4 images each of 150 ms duration: the initial fixation cross was followed by the first stimulus, another cross, then the second stimulus (figure 2). At the end of each trial the screen was set to mid-grey pending the observer's response.

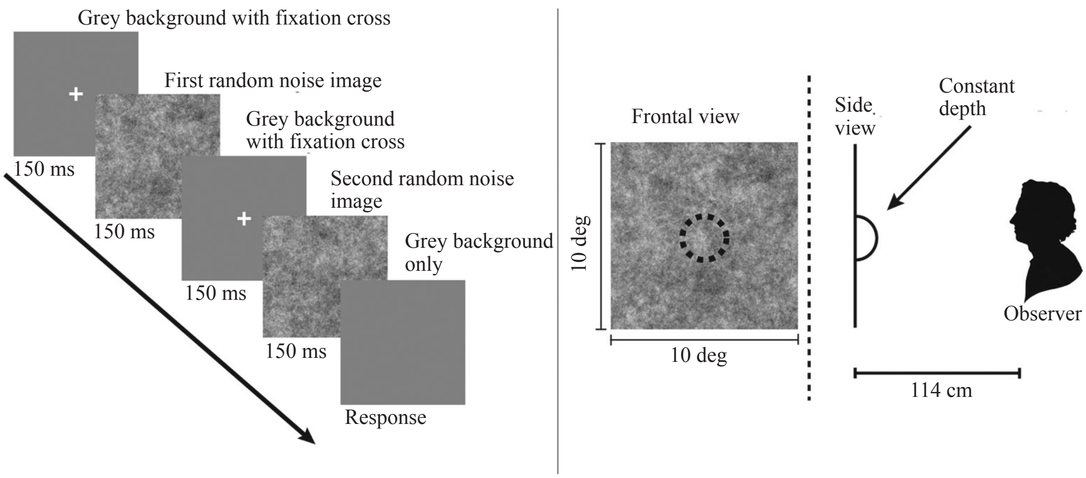

Figure 2. Left: experimental timeline. Right: frontal view showing relative size of small bump targets and side-view sketch shown as part of instructions.

As no signal was ever presented, we extended great caution when giving instructions because they represent the only information the participant has as to the nature of the target. At the start of the first session in any given condition observers were shown a side-view cross-sectional sketch of the target without any lighting information. They were asked to imagine what this shape might look like in a frontal view. At no point did observers see an example shaded target in frontal view. They were also instructed that the target, within each condition (session), had constant shape, size, and location and that—with the exception of the flat disk—they were always convex objects coming out of the screen. They were asked to fixate the centre of the screen throughout the experiment

We tested four target conditions: (i) small bumps, where the imagined target was a raised bump with a diameter of 5 cm (2.5 deg); (ii) big bumps, where the target was a raised bump with a diameter of 8 cm (4 deg); (iii) white disk, where the target was a flat white disk with a diameter of 5 cm (2.5 deg); and (iv) cylinder, where the target was a 5 cm high (2.5 deg) and 8 cm wide (4 deg) semicylinder.

2.4 *Analysis*

The classic CI technique typically uses weak target stimuli embedded in strong noise fields and asks observers to detect some target feature. The noise samples (excluding any target) are then accumulated according to the participant's response on each trial. There are four possible stimulus-response categories for each trial. Target-present trials can lead to 'hits', where the observer gives a positive response, or 'misses' in case of negative responses. Target-absent trails can lead to 'false alarms', if the observer gives a positive response, or 'correct rejections'. It has been shown (Ahumada and Beard 1996) that the correlation between the luminance of each pixel and the observer's response can be found by accumulating the CI across trials based on the four stimulus-response categories: CI = (hit+false alarm)−(miss+correct rejection). One simply adds or subtracts the relevant images on a pixelwise basis. As we had no targets in our experiments, there were only two stimulus-response pairings and templates were computed by accumulating all the images that led to positive responses ('yes images') into one pool and all the images that led to negative responses ('no images') into another pool. The CI is then given by the equation:

$$\text{CI} = \text{yes images} - \text{no images}.$$

As with the classical method, the resulting image represents the template of information used by the subject to perform the task. If the observer does not use a systematic approach, the resulting template should have the same properties as averaged noise. If not, the template indicates the presence of structures that underlie the perception of a target; and, more importantly, given the lack of a target in our stimuli, the observers' impression of what the target should look like given the instructions. We further adapted the CI method to use a 2IFC—this is similar to Abbey and Eckstein's (2000) two-alternative forced-choice method, but here we have two temporal intervals—such that on each trial the participant saw two similar noise samples (no target stimulus) and indicated which looked most like the imagined target. Thus each trial contributed an image to each of the positive and negative pools.

Templates comprised 512 × 512 images. We fit a Gaussian curve to the central column and central row of each template (see table 1). The peak of the best-fit Gaussians gives the location of any highlight in the template, while standard deviations indicate the size of such features. We estimated people's lighting prior as the difference between the centre of the screen and the peak of the best-fit Gaussian.

Table 1. Gaussian fit data for each participant and condition. Mean peak is an index of their lighting prior bias (as 256 is the centre of the image a smaller number would result in light-from-above preference while a bigger number would subtend a light-from-below bias); R^2 is an index to the goodness of fit that highlights if the observer performed the task by using a consistent strategy across trials; amplitude and variance are relative to the fitted curve; and sessions indicates the number of sets observers needed to produce a discernable template.

Participant's initials	Standard deviation	Mean (peak position)	Amplitude	R^2	Sessions
Small bump					
TY	30.81	215.9	58.15	0.6103	5
PS	43.97	244.9	87.8	0.9083	4
EL	42.29	256.2	156.3	0.9091	6
KG	43.69	266.4	80.78	0.8247	8
HB	40.31	254.2	232.1	0.9187	7
JH	41.88	258.0	99.41	0.7601	10
NK	46.27	242.9	52.32	0.6148	7
Big bumps					
TY	41.77	217.4	56.49	0.7594	5
PS	42.40	228.5	97.21	0.8356	4
EL	51.03	226.7	55.17	0.6456	6
KG	45.75	263.0	36.43	0.4056	6
White disk					
JH	44.59	243.9	159.1	0.7380	4
HB	46.88	256.8	105.9	0.7438	4
Cylinder					
HB	39.77	254.5	95.99	0.8262	4

3 Results

Observers produced clear target templates for each condition, despite the fact that no signal was ever presented. Most claimed that the task was very hard and felt they were guessing in early sessions, but after a few sessions they felt more comfortable with the task and claimed to have found a strategy for it. No one ever doubted of the presence of the target. Figure 3 shows example templates for each of the conditions tested, vertical sections, and best-fit Gaussians for each participant. As we used $1/f$ noise, the resultant CIs are blurred estimates of the observers' templates (Abbey and Eckstein 2000). Gaussian fit results are reported in table 1; R^2 values are included to give an index of goodness of fit. Only half of the small bump templates revealed a light-from-above prior. Figure 4a plots distance from the centre of the screen in terms of degrees of visual angle, for each observer in both small bump and large bump conditions. Figure 4b plots mean offsets in the small and large bump conditions. The mean offsets in the small bump condition was significantly above the midline on a one-tailed t-test ($t_6 = 2.164$, $p = 0.037$), as was the mean offset in the large bump condition ($t_3 = 3.844$, $p = 0.031$). Observers tended to keep the same preference for light from above or below despite changes in the target size. Although highlights were on average further from the centre in the large bump condition than in the small bump condition, this difference was not significant.

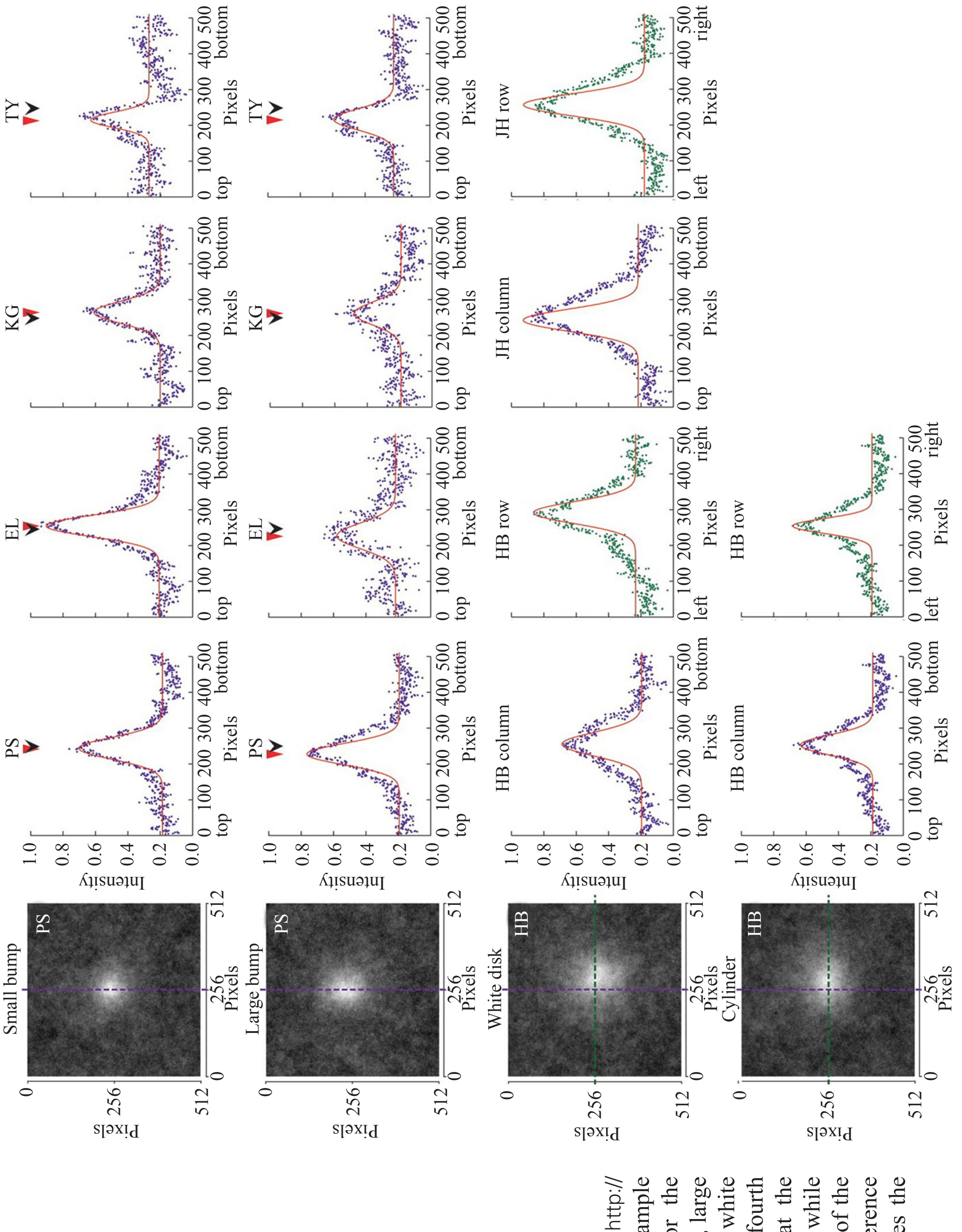

Figure 3. [In colour online, see http://dx.doi.org/10.1068/p75127] Example templates and cross-sections for the small bump condition (first row), large bump condition (second row), white disk (third row), and cylinder (fourth row). The black arrow points at the centre of the image (column 256), while the red arrow points at the peak of the Gaussian fit. The sign of the difference between the two arrows indicates the direction of the lighting prior.

Qualitatively, templates for the cylinder condition closely followed the small bump templates along the vertical axis but had a much wider profile on the horizontal axis. This reflects the elongated highlight one would expect from a cylindrical surface as compared with a spherical one. In the white disk condition, where the imagined object should not convey any lighting bias, template peaks were not offset from the centre of the image and templates appear somewhat broader and flatter (that is, more disk like) than the 'bump' templates. These observations confirm the task-dependent nature of our templates and thus the validity of the peak offsets measured in the bump conditions. The good Gaussian fits to the bump templates suggest that people were expecting a concentrated highlight in the bump condition, but the lack of consistency between observers in the location of the highlight (see figure 4a) suggests that the lighting-from-above prior is not universal. These individual differences suggest that the lighting-from-above prior may be less robust and less prevalent than previously thought.

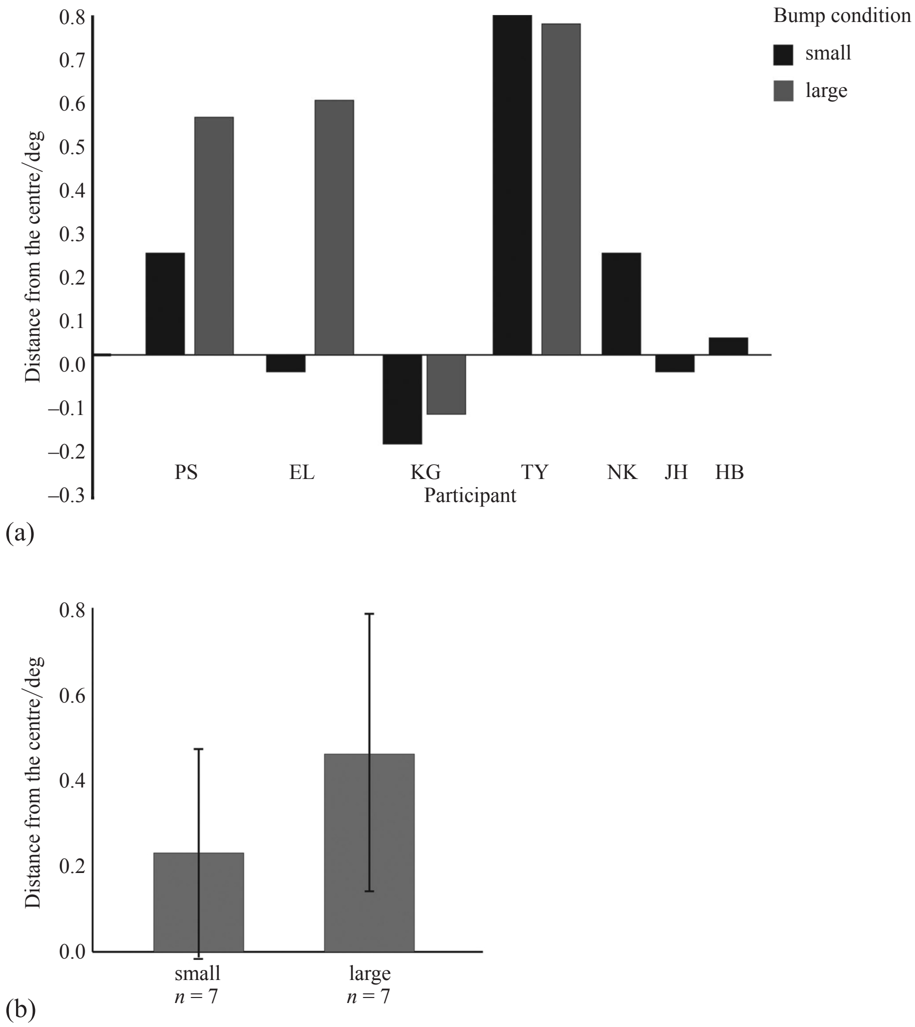

Figure 4. (a) All participants offset for both small and large bump conditions. (b) Lighting prior bar graph for small and large bumps. The distance from the centre is reported in terms of magnitude (no direction). Bars represent confidence intervals at 95%.

4 Discussion

Assumptions made about the position of the light source play a crucial role in the perception of ambiguous stimuli (Brewster 1826; Ramachandran 1988; Rittenhouse 1786; Todd and Mingolla 1983). In shape from shading the use of ambiguous images has always been considered the best way to test observers' prior assumptions (Gerardin et al 2007, 2010; Kleffner and Ramachandran 1992; Mamassian and Goutcher 2001; Morgenstern et al 2011; Ramachandran 1988; Sun and Perona 1998). Previous studies have found that a very high proportion of observers have a bias for seeing lighting from above at both long (Adams et al 2004; Ramachandran 1988; Sun and Perona 1998) and short (Mamassian and Goutcher 2001) presentation time. However, these studies presented ambiguous stimuli which nonetheless depicted shaded images and may therefore have promoted the assumptions of a directional light source over, say, a diffuse interpretation. We present an alternative method using noise-only stimuli which has less potential for biasing observer responses in favour of a particular lighting type. A consequence of testing with no stimulus, as we have done, might be to reveal the latent robustness of the lighting-from-above prior. The large variations we have found across participants suggest in fact that the light-from-above prior is fragile even in cases where a clear highlight is found. However, such highlights confirm that the observer is expecting to see a shaded object.

In summary, the adoption of a lighting prior is a logical strategy for the visual system to solve ambiguities in shaded stimuli since the light-from-above assumption has ecological validity. Nevertheless, the large variation found here reveals its influence to be somewhat less strong than might be expected from the literature. Our results therefore contribute to the growing evidence that the light-from-above prior is weak and dependent on stimulus features.

Acknowledgments. This work was funded by EPSRC Grants EP/F026269/1 to AJS.

References

Abbey C K, Eckstein, M P, 2000 "Estimates of human-observer templates for a simple detection task in correlated noise" *Proceedings of SPIE* Volume 3981 *Medical Imaging 2000: Image Perception and Performance* Ed. E A Krupinski (Bellingham, WA: SPIE) pp 70–77, doi:10.1117/12.383092

Adams W J, Graf E W, Ernst M O, 2004 "Experience can change the 'light-from-above' prior" *Nature Neuroscience* **7** 1057–1058, doi:10.1038/nn1312

Ahumada A J, 1996 "Perceptual classification images from Vernier acuity masked by noise" *Perception* **25** ECVP Supplement, 18 (abstract)

Ahumada A J, Beard B L, 1996 "Object detection in a noisy scene", in *Proceedings of SPIE* Volume 2657 *Human Vision and Electronic Imaging* Eds B E Rogowitz, J P Allebach (Bellingham, WA: SPIE) pp 190–199, doi:10.1117/12.238715

Brewster D, 1826 "On the optical illusion of the conversion of cameos into intaglios, and intaglios into cameos, with an account of other anologous phenomena" *Edinburgh Journal of Science* **4** 99–108

Gerardin P, de Montalembert M, Mamassian P, 2007 "Shape from shading: New perspectives from the Polo Mint stimulus" *Journal of Vision* **7**(11):13, 1–11

Gerardin P, Kourtzi Z, Mamassian P, 2010 "Prior knowledge of illumination for 3D perception in the human brain" *Proceedings of the National Academy of Sciences of the USA* **107** 16309–16314, doi:10.1073/pnas.1006285107

Gold J M, Murray R F, Bennett P J, Sekuler A B, 2000 "Deriving behavioural receptive fields for visually completed contours" *Current Biology* **10** 663–666, doi:10.1016/s0960-9822(00)00523-6

Gosselin F, Schyns P G, 2003 "Superstitious perceptions reveal properties of internal representations" *Psychological Science* **14** 505–509, doi:10.1111/1467-9280.03452

Kleffner D A, Ramachandran V S, 1992 "On the perception of shape from shading" *Perception & Psychophysics* **52** 18–36, doi:10.3758/bf03206757

Langer M S, Bülthoff H H, 2000 "Depth discrimination from shading under diffuse lighting" *Perception* **29** 649–660, doi:10.1068/p3060

Liu B X, Todd J T, 2004 "Perceptual biases in the interpretation of 3D shape from shading" *Vision Research* **44** 2135–2145, doi:10.1016/j.visres.2004.03.024

Mamassian P, Goutcher R, 2001 "Prior knowledge on the illumination position" *Cognition* **81** B1–B9, doi:10.1016/s0010-0277(01)00116-0

Morgenstern Y, Murray R F, Harris L R, 2011 "The human visual system's assumption that light comes from above is weak" *Proceedings of the National Academy of Sciences of the United States of America* **108** 12551–12553, doi:10.1073/pnas.110794108

Murray R F, 2011 "Classification images: A review" *Journal of Vision* **11**(5):2, 1–25, doi:10.1167/11.5.2

Ramachandran V S, 1988 "Perception of shape from shading" *Nature* **331**(6152) 163–166, doi:10.1038/331163a0

Rittenhouse D, 1786 "Explanation of an optical deception" *Transactions of the American Philosophical Society* **2** 37–42

Schofield A J, Rock P B, Georgeson M A, 2011 "Sun and sky: Does human vision assume a mixture of point and diffuse illumination when interpreting shape-from-shading? *Vision Research* **51** 2317–2330, doi:10.1016/visres.2011.09.004

Sun J, Perona P, 1998 "Where is the sun?" *Nature Neuroscience* **1** 183–184, doi:10.1038/630

Todd J T, 2004 "The visual perception of 3D shape" *Trends in Cognitive Sciences* **8** 115–121, doi:10.1016/j.tics.2004.01.006

Todd J T, Mingolla E, 1983 "Perception of surface curvature and direction of the illumination from patterns of shading" *Journal of Experimental Psychology: Human Perception and Performance* **9** 583–595, doi:10.1037/0096-1523.9.4.583

Tyler C W, 1997 "Diffuse illumination as a default assumption for shape-from-shading in the absence of shadow", in *Proceedings of SPIE* Volume 3016 *Human Vision and Electronic Imaging II* Eds B E Rogowitz, T N Pappas (Bellingham, WA: SPIE) pp 346–351, doi:10.1117/12.274530

Watson A B, 1998 "Multi-category classification: Template models and classification images" *Investigative Ophthalmology & Visual Science* **39** S912