# Sun and sky: Does human vision assume a mixture of point and diffuse illumination when interpreting shape-from-shading?

Schofield, Andrew; Rock, PB; Georgeson, MA

*DOI:*
10.1016/j.visres.2011.09.004



*Document Version*
Peer reviewed version

*Citation for published version (Harvard):*
Schofield, A, Rock, PB & Georgeson, MA 2011, 'Sun and sky: Does human vision assume a mixture of point and diffuse illumination when interpreting shape-from-shading?', Vision Research, vol. 51, no. 21-22, pp. 2317-2330. https://doi.org/10.1016/j.visres.2011.09.004

# Sun and Sky: Does human vision assume a mixture of point and diffuse illumination when interpreting shape-from-shading?

Andrew J Schofield[1], Paul B Rock[1], Mark A Georgeson[2]

1. School of Psychology, University of Birmingham, Edgbaston, Birmingham, UK, B15 2TT
2. School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham, UK, B4 7ET UK

AJS: a.j.schofield@bham.ac.uk,
MAG: m.a.georgeson@aston.ac.uk

Corresponding author: Andrew. J. Schofield.
Email: a.j.schofield@bham.ac.uk.



Supplementary file appended at end of document.

**Abstract**

People readily perceive smooth luminance variations as being due to the shading produced by undulations of a 3-D surface (shape-from-shading). In doing so, the visual system must simultaneously estimate the shape of the surface and the nature of the illumination. Remarkably, shape-from-shading operates even when both these properties are unknown and neither can be estimated directly from the image. In such circumstances humans are thought to adopt a default illumination model. A widely held view is that the default illuminant is a point source located above the observer's head. However, some have argued instead that the default illuminant is a diffuse source. We now present evidence that humans may adopt a flexible illumination model that includes both diffuse and point source elements. Our models estimates a direction for the point source and then weights the contribution of this source according to a bias function. For most people the preferred illuminant direction is overhead with a strong diffuse component.

Keywords: shading, illumination, lighting-from-above, dark-is-deep.

## 1. Introduction

### *1.1 Background*

It is well known that humans can discern the shape of a surface from the pattern of shading produced when it is illuminated – shape-from-shading – even when there are no other cues to shape in the image (Christou & Koenderink, 1997; Erens, Kappers & Koenderink, 1993; Kleffner & Ramachandran, 1992, Langer & Bülthoff, 2000; Ramachandran, 1988; Todd & Mingolla, 1983; Tyler, 1998). Note however that shape-from-shading is not always veridical (Pentland 1988; Zhang, Tsai, Cryer, & Shah, 1999). To interpret shape-from-shading we must simultaneously estimate the shape of the surface, its reflectance properties, and the nature and direction of the illuminant – a task which is inherently ambiguous (D'Zmura, 1991; Belhumeur, Kriegman, & Yuille, 1999). Nonetheless a number of different cues enable humans to estimate the direction of illumination for a scene (Cavanagh & Leclerc, 1989; Erens et al., 1993; Gerhard & Maloney, 2010; Koenderink, van Doorn, Kappers, te Pas, & Pont, 2003; Koenderink, van Doorn, & Pont, 2004; Koenderink, van Doorn, & Pont, 2007; Liu & Todd, 2004; Norman, Todd, & Orban, 2004; Pentland, 1982; Todd & Mingolla, 1983) and although such estimates are not always accurate when the scene is well articulated we can estimate the light field with considerable accuracy (Koenderink, Pont, van Doorn, Kappers & Todd, 2007). In the absence of cues to lighting we may assume a default light source. There is debate, however, about the nature of the default light source. Several studies have suggested that humans assume a single spatially limited (point) light source located approximately overhead (Adams, Graf & Ernst, 2004; Mamassian & Goutcher, 2001; Ramachandram, 1992; Sun & Perona, 1998): lighting-from-above. In contrast, Langer & Bülthoff (2000) showed that humans can, if required, interpret shape-from-shading to be consistent with a diffuse, multidirectional light source. Tyler (1998) argues that diffuse illumination is the primary default assumption for highly reduced scenes.

The lighting-from-above assumption seems to explain a range of illusions – known collectively as the crater illusion – where, in monocular viewing, perceived surface shape flips from convex to concave when the image is rotated through 180° (Brewster, 1826; Hess, 1950; Ramachandran, 1992; Rittenhouse, 1786; von Fieandt, 1949). Lighting-from-above makes clear predictions about the relationship between shape and luminance. For a Lambertian surface, luminance at any point will be proportional to the cosine of the angle between the surface normal and the line joining the point to the light source. Parts of the surface that point towards the light source will have the highest luminance.

Diffuse illumination (such as used by Langer & Bülthoff, 2000) represents the situation on a cloudy day where a horizontal surface is illuminated about equally from all parts of a

hemispherical ‘sky’ this being (for surfaces) equivalent to a fully spherical illumination field or Ganzfeld (through this paper we use the terms diffuse or fully diffuse to mean this type of lighting). Diffuse illumination also leads to clear predictions about the relationship between shape and luminance. The luminance at any point on a diffusely illuminated Lambertian surface is approximately proportional to the size of the aperture formed by the rest of the surface (Langer & Zucker, 1997; Stewart & Langer, 1997; Tyler, 1998). Points down a slope, in a pit or in a ravine 'see' less of the sky and are hence dark (the dark-is-deep rule). However, at the very bottom of a ravine or pit the surface points directly towards the un-obscured sky producing a small localized luminance peak (see Langer & Bülthoff, 2000).

Although both the lighting-from-above and diffuse illumination models have some ecological validity, neither correspond well to everyday lighting conditions. Humans are generally immersed in an illumination field that is highly diffuse but biased towards the sky because of the location of the sun (or room lights) and the relatively low reflectance values of ground-level objects. For real, outdoor situations the illumination reaching a point from any given direction decreases monotonically with decreasing elevation except for a local dip around horizontal (Dror, Willsky & Adelson, 2004; Teller, Antone, Bosse, Coorge, Jethwa, & Masters, (2001); see also Mury, Pont & Koenderink, 2009). It is likely then that people assume a default illumination model that resembles everyday experience and that therefore, when the nature of the illuminant is uncertain, they assume that objects are lit by a light field that is quite diffuse but with a directional component.. We test this hypothesis here.

*1.2 Choice of stimuli*

Langer and Bülthoff (2000) presented observers with images of complex undulating surfaces rendered under either point source or diffuse lighting. The resulting depth judgements show that humans are able to switch between point and diffuse light interpretations depending on cues in the stimulus presented. This suggests that the default illumination model assumed by human vision can only be exposed when the stimuli are ambiguous with respect to illumination. We therefore avoid complex rendered stimuli and present instead simple stimuli which we show are likely to be ambiguous with respect to illumination. In this we follow the lead of Sun & Perona (1998) and Mamassian & Goucher (2001) who presented stimuli where the direction of the illumination was ambiguous. In our case, however, it is the nature of the illumination (diffuse vs point) that is uncertain. We also need test stimuli that are expected to produce quantitatively different results depending on the nature of the assumed illumination.

People tend to perceive sinusoidal shading patterns[1] as sinusoidally undulating surfaces (see Pentland, 1988 and Schofield, Hesse, Rock & Georgeson, 2006; we also present control data in the supplementary file to further support this claim using stimuli from our main experiments) despite the fact that such surfaces do not always give rise to sinusoidal luminance profiles when shaded (see supplementary Figures S1 & S2). However, the analysis presented below and in the supplementary file shows that sinusoidal undulations do give rise to approximately sinusoidal shading profiles under point source lighting for a range of surface orientation; Figure 1 shows examples where this is the case.

**Figure 1 about here (double column)**

We see from Fig 1 that point sourse lighting produces either approximately sinusoidal shading profiles with luminance peaks offset from the physical surface peaks by ¼ wavelength, double crested peaks centred on the ¼ wavelength offset, or a frequency doubled signal. The ¼ wavelength offset is counter intuitive and we now show that this offset does not vary with either the surface or lighting conditions so long at the shading profile has a single peak. Following Pentland (1988) we approximate the luminance at any point on a Lambertian surface with the following equation:

$$L \approx \cos\theta + p\cos\tau\sin\theta + q\sin\tau\sin\theta - \cos\theta(p^2 - q^2)/2 \qquad \text{eqn1}$$

where $\theta$ is the slant of the light source (the angle that the illuminant vector makes with the *z*-axiz) $\tau$ is the tilt of the light source (the angle between the *x*-axis and the projection of the light source vector onto the surface plane, *p* is the partial derivative of the surface with respect to *x* and *q* its partial derivative with respect to *y*. Let the surface $z = a\cos(\omega x)$, where $\omega$ is the undulation frequency and *a* is the surface amplitude; hence $p = -a\omega\sin(\omega x)$ and $q$=0. We further redefine the lighting angles in terms of the elevation angle $(e = \pi/2 - \theta)$ between the light vector and the image plane and direction angle $(d = \tau \mp \pi 2$ being the angle between the projection of the light vector into the surface plane and the y-axis (vertical) we define positive changes in *d* as clockwise rotations (see Fig 1B for a diagram of the lighting geometry). Equation 1 can thus be rewritten as:

$$L \approx \cos\left(\frac{\pi}{2} - e\right) - a\omega\sin(\omega x)\cos\left(\frac{\pi}{2} - d\right)\sin\left(\frac{\pi}{2} - e\right) - \cos\left(\frac{\pi}{2} - e\right)\left(a^2\omega^2 sin^2(\omega x)\right)/2 \qquad \text{eqn2}$$

[1] Stimuli where luminance is a sinusoidal function of position in the image.

In order to find the locations of peaks and troughs in *L* we need to differentiate eqn 2 with respect to *x*:

$$\frac{dL}{dx} \approx -a\omega^2 \cos(\omega x) \cos\left(d + \frac{\pi}{2}\right) \sin\left(\frac{\pi}{2} - e\right) - \cos\left(\frac{\pi}{2} - e\right) a^2 \omega^3 \sin(2\omega x)/2 \qquad \text{eqn3}$$

This has two components one with extreama located at $\pi/2$ and $3\pi/2$ (frequency = $\omega$; offset = $\pi/2$) and the other with extreama at 0, $\pi/2$ , $\pi$, and $3\pi/2$ (frequency = $2\omega$). The locations of these extreama do not change with surface amplitude or frequency nor with lighting direction however the ratio of the two components does change introducing double crested peaks and ultimately frequency doubling for some lighting conditions in a manner that also depends of *a* and $\omega$ (larger values favour double peaks). Aside from cases where the frequency doubled term dominates completely luminance will always have a positive lobe at either $\pi/2$, or $3\pi/2$ (1/4 wavelength offset). Double peaks occur by virtue of local minima at these locations. Thus we can identify double crested peaks by examining the extreama at $\pi/2$, and $3\pi/2$; if both are minima then the peak is double crested. This in turn can be determined from the second derivative of luminance:

$$\frac{d^2L}{dx^2} \approx a\omega^3 \sin(\omega x) \cos\left(d + \frac{\pi}{2}\right) \sin\left(\frac{\pi}{2} - e\right) - \cos\left(\frac{\pi}{2} - e\right) a^2 \omega^4 \cos(2\omega x) \qquad \text{eqn4}$$

setting *x*=$\pi/2$ or $3\pi/2$. We can thus in principle find the lower limit of tilt giving single peaks for each combination of slant, amplitude and frequency. For the amplitude : wavelength ratio used in the current study (0.12) this lower limit is depicted by the border of the inner black lozenge in supplementary Fig S2b. It is clear that, in the absence of double peaks, luminance will always peak at an offset of ¼ wavelength from the physical surface peak and that when double crested peaks do occur they will always lie either side of lobe centred on ¼ wavelength offset. When full frequency doubling occurs luminance peaks will always occur at the peaks and troughs of the surface. Finally frequency doubling will always occur when the elevation of the light is $\pi/2$ (frontal lighting) or when the lighting direction is either zero or $\pi$ (lighting from above or below)

Diffuse lighting where the surface is illuminated from all directions will produce approximately sinusoidal shading regardless of the surface orientation (Figure 2). Comparing Figures 1 and 2 we note that point source illumination produces either a ¼ wavelength offset between

surfaces peaks and luminance peaks or results in frequency doubling[2] , whereas diffuse lighting produces neither an offset nor frequency doubling.

Figure 2 about here – Single column

It follows from the superposition rule that a mixture of diffuse and point source lighting will produce waveforms that are approximately sinusoidal but with luminance peaks that are offset from the physical surface peaks by some amount between 0 and ¼ wavelength. The addition of two sine waves with the same frequency but different amplitudes and phases being a sine wave with the same frequency but intermediate phase. This will hold so long as the point source term is not dominated by frequency doubling. There will also be a localised peak at the surface troughs due to the diffuse lighting component. Figure 3 plots the luminance profiles for oriented surfaces under mixed illumination in the format of Fig 1. The offset between the fundamental and the physical surface clearly changes with the physical orientation of the surface; as indicated in Fig 3. We note that even for vertically oriented surfaces lit from above the shading profile is dominated by the fundamental not the frequency doubled component.

Figure 3 about here double column

At least in terms of the generative models outlined above and in the supplimentary file the relationship between surface profiles and shading is different for the three types of lighting even though sinusoidal shading is a reasonable approximation in all cases. Point source lighting produces ¼ wavelength offsets; diffuse lighting - zero offset; and mixed illumination offsets that vary with surface orientation. We thus propose sinusoidal shading patterns as a diagnostic stimulus for the default illumination model used in human shape-from-shading. If people were apply the inverse of the appropriate generative model (at least approximately) to estimate shape, we could determine which model had been adopted by observing the offsets between luminance and perceived surface peaks (inter-peak offset) and the tendency towards frequency halving at some orientations (undoing the doubling found in the point source case). This assumption is central to our method so we expand on it below.

Point source assumption: how might a ¼ wavelength inter-peak offset arise in human vision? If people were to assume point source lighting then the peaks of the perceived surface should (in general) be shifted away from the luminance peaks. The linear relationship noted

[2] Under conditions where approximately sinusoidal shading is achieved, see supplementary file.

by Pentland (1988) will be valid if the direction/elevation of the light source is oblique (or assumed to be oblique). In Pentland's (1988) model for human shape-from-shading, luminance components are subject to a 90° phase shift in the frequency domain. Thus for sinusoidal shading under Pentland's model, perceived surface peaks will be offset by ¼ wavelength from luminance peaks: nicely undoing the offset produced by point-source shading in the first place (see Figures 1 & supplementary file). Alternatively recovering shape-from-shading is sometimes characterized as an integration process in which perceived surface gradient is proportional to luminance. Integration would also produce a ¼ wavelength inter-peak offset for sinusoidal shading, although the presence of bounding contours will alter the integration process by setting its boundary conditions. Thus the generative model, Pentland's model, and integration models all support the notion that sinusoidal luminance profiles should be seen as sinusoidal surfaces with a ¼ wavelength offset between perceived surface peaks and luminance peaks *if* the observer assumes a point light source.

In cases where a point source illuminant is aligned with the surface undulations (eg. upper trace in Figure 1) the shading profile is dominated by a quadratic component (frequency doubling). If people were to allow for such quadratic shading we would expect them to see stimuli aligned with their assumed point source direction as surfaces undulating at half the spatial frequency of the shading pattern. However, Pentland (1988) has shown that people do not allow for quadratic shading when interpreting shape-from-shading although, as we outline in Section 1.4, this alone does not mean that people do not assume point source lighting in this special case.

Diffuse source assumption: how might a zero inter-peak offset arise in human vision? Langer and Bülthoff (2000) found that when surfaces are rendered under diffuse illumination humans adopt the dark-is-deep rule whereby peaks in the perceived surface align with luminance peaks. The strict application of the dark-is-deep rule would predict a small localized peak in the perceived surface in the bottom of valleys due to the local peak in luminance at such points, but Langer & Bülthoff (2000) found that people do not perceive peaks at these locations. Rather, their data were best characterized by a model in which the luminance profile associated with the surface was first blurred, attenuating small local peaks, and then interpreted according to the dark-is-deep rule. Note that such blurring could render the luminance profile of Figure 2b identical to a sinusoidal profile and that if the stimulus is itself sinusoidal, blurring, by say a Gaussian filter, will only alter the amplitude of the signal: not its shape or position . If people were to assume diffuse lighting Langer and Bulthoff's

(2000) blur+dark-is-deep model would predict a sinusoidal surface interpretation but with no offset between perceived luminance peaks and surface peaks.

Mixed source assumption: how might intermediate inter-peak offsets arise in human vision? There is little support in the literature for this case. However, noting that shading profiles under mixed illumination tend to be quite irregular, suitable offsets could be achieved in one of two ways: first by reversing the image generation process under a mixed lighting assumption; and second a combination of the blur+dark-is-deep rule and Pentlands (1988) model with a stimulus dependent weighting between the two. We show later that for sinusoidal shading patterns these two alternatives make identical predictions.

Thus we are confident that people are, in principle, capable of interpreting our stimuli according to either a point or diffuse lighting assumption; and we can see a route by which a mixed illumination assumption might be implemented. The question is which assumption dominates.

### 1.3 Bas-Relief ambiguity

Point-source lighting of surfaces produces ambiguous luminance profiles due to the generalised Bas-Relief (GBR) ambiguity and the related convex/concave ambiguity (Belhumeur, et al., 1999; see also D'Zmura, 1991). Any shaded surface can be modified by a GBR transformation which when coupled with a suitably transformed lighting and albedo profiles will produce the same luminance profile as the original surface and lighting conditions. Humans are thus unable to make good judgements about (for example) the amount of relief applied two sculptures. This ambiguity has some relevance to shape-from-shading studies in general. Of more critical interest here however is the convex/concave ambiguity in which a convex surface lit from below looks identical to a concave surface lit from above. Prior assumptions for convexity and lighting-from above serve to stabilise this percept and prevent perceptual flipping (Liu & Todd, 2004; Sun & Perona, 1998; and Mamassian & Goucher, 2001). However, the convexity prior will not apply to sinewave shading which appears corrugated (both convex and concave) and lighting from above will only function for near horizontal stimuli. There is a strong chance then that the perceived position of peaks in near vertical stimuli will flip between two possible positions. We explicitly test for this.

### *1.4 Experimental predictions*

Our main aim was to assess the default illumination model used by observers to interpret the perceived shape of simple shading patterns in the absence of other cues to surface shape.

We also wanted to avoid explicit cues to the nature of the light source. Sinusoidal luminance patterns can approximate the shading obtained under both point-source illuminants (Figure 1) and, to a lesser extent, diffuse illumination (Figure 2) while avoiding the above confounding factors; we therefore used sinusoidal gratings as our shading stimuli (we did not use rendered stimuli; see Section 2 for further justification). Observers were free to adopt any lighting hypothesis in order to 'make sense of' the stimuli. We presume that observers may have a preference for lighting-from-above (Adams, etal., 2004; Mamassian & Goutcher, 2001; Ramachandram, 1992; Sun & Perona, 1998) when adopting a point-source hypothesis, but this is by no means fundamental to the experiment.

An important diagnostic case occurs when sinusoidal shading patterns align with the observers preferred lighting direction for point source illumination. Given that most observers prefer lighting from above (Mamassian and Goucher, 2001) this special case most often corresponds to a vertical sinewave. We have no reason to suppose that this is any less common a visual experience than any other sinewave. We test five predictions for this critical case. (1) If people perceive such surfaces to be lit from their preferred direction by a point source, and have at least an implicit model of the physics of shading under such conditions, then they would perceive such a surface to have half the frequency of the luminance profile (undoing the quadratic shading or frequency doubling seen for vertical surfaces in Figure 1). (2) Alternatively, people might perceive such surfaces as lit by a point-source but alter their estimate of the direction of this source consistently to one side or the other. If this were so they would perceive a surface with the same frequency as the shading and would retain the inter-peak offset expected at other orientations. (3) We might, however, expect such an interpretation to be bi-stable owing to the convex-concave ambiguity (section 1.3) which is most intrusive when shading gradients are orthogonal to the observer's preferred light source (Sun and Perona, 1998). Such bi-stability would result in the inter-peak offset flipping between two locations either side of zero. Here the average offset would fall to zero but the distribution of offsets would become bi-modal. (4) If (as suggested by Langer & Bülthoff, 2000) people can switch between point- and diffuse lighting interpretations depending on stimulus cues, they might prefer a diffuse model for (close to) vertical stimuli. If so they should shift from using the luminance=gradient 'rule' to the 'dark-is-deep' rule and inter-peak offsets will vary accordingly. (5) Finally, shape-from-shading may fail for some stimuli. Specifically, sinusoidal stimuli may fail to elicit a depth percept at some orientations, causing people to perceive surfaces as flat at these orientations, thus degrading estimates of inter-peak offset.

## 2. Experiment 1. Inter-peak offsets

The purpose of this experiment was to measure the spatial offset between the luminance peaks of a shading pattern and the associated peaks of the perceived surface. In particular we asked how this inter-peak offset varies with stimulus orientation in the frontal plane. We used a haptic matching task in which observers adjusted the position of a haptically defined sinusoidal surface to match that of a visually perceived surface. In contrast to most studies of shape-from-shading, but in common with Pentland (1988) and Kingdom (2003), our stimuli (see Figure 3) were sinusoidal gratings imposed on iso-tropic textures. They were not rendered surfaces. The textures help to articulate the shading cue but introduce no depth cues in themsleves. We used these stimuli because they give the observer freedom to interpret the shading cue in the absence of other cues to shape or overt cues to the nature of the light-source. We justify this as follows: (1) Our stimuli contain no geometric cues, either in the form of distortions in the texture or bounding contours, that might otherwise bias the shape-from-shading process. (2) They do not include any sharp luminance edges that could be associated with shadows nor do they contain double-crested peaks, and so they do not promote a point-source lighting interpretation. (3) They do not contain mini-peaks between each luminance peak and therefore do not promote a diffuse lighting interpretation either. (4) As anisotropic stimuli they are largely uninformative about the direction of the light source (Koenderink, et al., 2007). Despite the fact that the visual stimuli were not produced by a graphical rendering of a model surface, observers readily perceived the stimuli as corrugated surfaces as we show in the supplementary file (Section S2).

**Figure 4 about here (2 column)**

By using highly under-constrained stimuli we hope to reveal internal observer biases. In particular our stimuli are mostly free from cues that might promote either a point or diffuse lighting interpretation (although they are a better approximation for point-source lighting: *cf* Figures 1 and 2). In this sense we differ from Langer & Bülthoff (2000) who used realistic rendering to bias observers to one or other light source type and Tyler (1998) whose radial sine waves could not be interpreted as lit by a single point-source.

Based on the results of Schofield et al. (2006; see also Pentland, 1988, and the control experiment in the supplementary file) we suppose that humans naturally perceive sinusoidal luminance profiles as sinusoidally corrugated surfaces. However, it is possible that humans adopt a very flexible approach to shape-from-shading, balancing a number of *a priori* constraints so as to perceive the combination of surface shape and illumination profile that is most likely to occur in real world situations. Therefore, given our overall aim of assessing the lighting model used by observers, we felt it important – at least in the first instance – to fix

the surface interpretation. Asking observers to match haptically defined sine waves to visual stimuli should enhance the impression that the visible surfaces were sinusoidal (Wijntjes, Volcic, Pont, Koenderink, & Kappers, 2009) thus leaving their internal lighting model as the only thing free to vary in order to 'make sense of' the stimuli presented.

*2.1 Method*

2.1.1 Procedure and stimulus details.

Observers adjusted the position of the undulations of a virtual haptic surface to match the perceived undulations in a visually presented stimulus. Visual stimuli (see Figure 3) consisted of an isotropic, binary visual noise texture (mean contrast=0.1) whose luminance values were multiplied by a sinusoidal profile ($1+c.\sin(2\pi fx)$; spatial frequency f=0.4 c/deg, luminance contrast c= 0.2) so as to emulate multiplicative shading in which the local mean luminance of the surface texture is modulated but not its local contrast. Such signals can be produced by adding a sinewave luminance modulation while modulating the local amplitude of the noise texture in phase with the luminance signal (LM+AM, see Schofield, et al., 2006 for details). The orientation of this shading pattern varied over the range 0-165° at 15° intervals. Note that we measured stimulus orientation clockwise from vertical but later (and in supplementary Fig S1) express positive increments in the direction of the illuminant as anti-clockwise rotations. We use this convention because, in terms of the shading pattern produced, a clockwise rotation of the stimulus is equivalent to an anti-clockwise shift in the direction of the illuminant.

The wavelength of the sine wave modulation was 25mm and its phase was randomized on a trial-by-trial basis. An orthogonal sinusoidal signal comprising both a luminance modulation and an anti-phase amplitude modulation (LM-AM, see Schofield et al, 2006) was added to each stimulus. This component was irrelevant to the current study but was included because the experiment was part of a larger study where observers' perception of the LM-AM component was relevant. We have previously shown that the LM-AM combination is seen as flat stripes in these plaid patterns and that the perception of the LM+AM component varies little with the presence of this extra cue (Schofield et al., 2006; Schofield, Rock, Sun, Jiang & Georgeson, 2010).

Haptic stimuli consisted of a virtual surface with sinusoidal undulations collocated with the visual stimulus and presented at the same orientation and spatial frequency as the shading signal. These stimuli were presented via a small force-feedback robot arm with a pen-like stylus. The arm provided physical resistance whenever the observer tried to move the stylus tip through the virtual surface. Observers held the stylus with their dominant hand and gently

stroked the virtual surface. The initial position of the surface relative to the visual stimulus varied at random from trial to trial. Surface amplitude was fixed at ±3 mm (amplitude = 0.12 of a wavelength). Three markers (not shown in Figure 3) were added to the visual stimulus: one at fixation and two at opposite edges of the stimulus positioned such that the alignment of the three markers indicated the direction in which observers should stroke the haptic surface in order to feel the undulations. The position of the stylus tip was marked by a small circle to provide visual feedback of the stylus location. Visual and haptic stimuli were generated on a PC computer and observers adjusted the position of the haptic stimulus using keys 4 and 6 on the computer's numeric keypad. Numbers placed next to the outer markers in the visual stimulus indicated which key to press to move the haptic surface towards each marker. The haptic surface could be moved in either 1.4 or 0.35 mm steps, toggled as required by pressing key 5. Observers heard a long tone after each 1.4 mm movement and a short tone after each 0.35 mm movement.

### 2.1.2 Equipment and calibration

Stimuli were presented in a modified ReachIN haptic workstation (ReachIN Technologies AB, Stockholm, Sweden). The visual stimuli were presented on a 17" Sony Trinitron CPD G200 CRT monitor mounted at an angle of 45° above a horizontal half-silvered mirror. Haptic stimuli were presented via a Phantom-Desktop (SensAble Technologies Inc, Woburn, MA) force feedback arm located beneath the mirror. Observers looked into the mirror at a downward angle and thus perceived the visual stimulus to be beneath the mirror and approximately perpendicular to the line of sight. The effective viewing distance was about 57 cm. Visual stimuli were calibrated against the monitor's gamma characteristic using look up tables in a BITS++ graphics interface (CRS Ltd, Rochester, UK) which also served to enhance the available grey level resolution to the equivalent of 14 bits. Values in the look up tables were determined by fitting a four-parameter monitor model (Brainard, Pelli, & Robson, 2002) to luminance readings recorded with a CRS ColourCal photometer.

### 2.1.3 Observers

The 15 observers had normal or corrected to normal vision and, with the exception of authors AJS & PR, were paid for their time and unaware of the purposes of the experiment. They each undertook at least five observations at each orientation. Six of these observers (from a pilot study) were not tested at orientations 15,75,105 & 165°. Observations were made in a darkened room so that even though the mirror was half-silvered the observers could not see their own hand through it. A hood was fitted to the monitor such that observers could not view the screen directly. Head position was not physically constrained, but the arrangement of the hood and the need to keep the haptic stimuli at a comfortable distance

for one's arm served to limit head position. Head orientation was not constrained but observers were told to keep their heads upright. Viewing was binocular and so the visual stimulus provided stereoscopic cues to flatness. However, we have previously shown (Schofield, et al., 2006) that a robust percept of shape-from-shading can be derived from such stimuli, and binocular presentation avoids the rivalry associated with monocular presentation for some observers.

*2.2 Analysis*

The final position of the haptic surface was recorded at the end of each trial as was its offset relative to the nearest luminance peak in the shading profile (see Figure 4). Positive offsets (expressed as proportions of a wavelength) indicated that the perceived surface peak was below the luminance peak (or to the right at 0°) for orientations from 0-165°. Data were analysed by first taking medians (not all distributions were normal) then extrapolating the recorded data to cover the full range of haptic directions from 0 to 360°. To do this we exploited the fact that the orientation of the visual stimuli repeated every 180° whereas offset direction repeats only every 360°. Hence, positive offsets in the range 180-345° would correspond to a perceived surface peak above (to the left at 180°) of the luminance peak. Thus the extrapolated data in the range 180-345° were set to the negative of those recorded over the range 0-165°. This extrapolation is relevant for the modelling in Section 5.

*2.3 Results and discussion*

Figure 5 plots offsets between luminance peaks and perceived surface peaks (inter-peak offsets) as a function of stimulus orientation for the nine observers who provided observations at all orientations. There were considerable individual differences in behaviour but strong common themes emerged. Inter-peak offsets varied with stimulus orientation and typically ranged between 0 and ¼ wavelengths. The majority of observers (10 out of 15) produced their largest offset at orientations close to horizontal (90 & 270°) and their smallest offset close to vertical (0°) as exemplified by observers HS, HW, AS, AJS, AO & RCL. Five observers produced their largest and smallest offsets at some other orientations (e.g. PJ & AT). Based on the models described later we define the orientation orthogonal to each person's maximum offset as their illuminant aligned orientation. This orientation generally corresponds to a zero-crossing in the model offset traces of Figure 5 and is the point at which the perceived ridges 'ran' towards the observer's preferred light source as estimated by the model. Most observers perceived surface peaks to be below and to the right of the luminance peaks (consistent with lighting from above the line of sight), but three placed their surface peaks above the luminance peaks (e.g. AT, consistent with lighting-from-below). Seven observers showed a smooth transition between their maximum and minimum offsets

(e.g. HW, HS, AT, AJS, AO) whereas the remainder had more abrupt transitions. For example, all of PJ's offsets were close to ¼ wavelength; none were close to zero. The maximum absolute offset for some observers was noticeably less than ¼ wavelength (eg, HS & AO).

**Figure 5 about here (single column)**

We were worried about possible contamination from the convex/concave ambiguity. The perceived surface may be more ambiguous at some orientations than others and flips in the positions of perceived peaks could reduce offsets. If this were the case we would expect standard deviations to increase with decreasing offsets and for orientation with low offsets to have bi-modal distributions. We calculated the coefficient of bimodality $\left[b = (1 + skewness(x)^2)/\left(kurtosis(x) + 3\left((n-1)^2/\left((n-2)(n-3)\right)\right)\right)\right]$ for each observer at each orientation where *n* is the number of observations and where kurtosis is defined as being zero for a normal distribution. We then correlated this metric with offset magnitudes. If the concave/convex ambiguity were a problem we would expect a *negative* correlation between *b* and offset magnitudes. The mean correlation was significantly different from zero on a one sample t-test but was positive ($\bar{r}$=0.2, *t*=3.6, *df*=14, *p*=.003) implying that offset distributions tend to be bimodal when median offsets are large not small. Thus the concave/convex ambiguity cannot have resulted in the reduced offsets recorded.

In the introduction we proposed inter-peak offsets as a means to assess the nature of people's assumed light source. Point source interpretations should lead to ¼ wavelength offsets, a diffuse lighting assumption will produce no inter-peak offset and a mixed lighting assumption predicts offsets that depend on orientation. While some participants perceive surface peak to be offset from luminance peaks by ¼ wavelength at some orientations no offset was found at other orientations and some participants never perceived an offset as large as ¼ wavelength. The similarity between perceived inter-peak offsets and the pattern of physical inter-peak offsets observed for mixed illumination (Fig 3 & Model A, Section 5.1) suggests that many people assume a mixed lighting model .These results support prediction 4 (Section 1.4).

**3. Experiment 2: Perceived depth magnitude does not vary with stimulus orientation**. We were concerned to ensure that the magnitude of the perceived undulations did not vary systematically with orientation, and that there was no association between inter-peak offset and perceived depth. In particular, we wanted to verify that participants did not see illuminant

aligned stimuli as flat, as such a result might imply a failure to perceive shape-from-shading at the given orientation.

*3.1 Method*

Seven observers (all naïve to the purpose of the experiment; six of whom had previously taken part in Experiment 1) were presented with visual stimuli identical to those of Experiment 1 and additional single oblique LM+AM stimuli (left side of Figure 3). They were asked to adjust the amplitude of a collocated haptic surface to match that of the visually perceived undulations. Haptic stimuli were aligned with the LM+AM components of plaid stimuli and the offset between the haptic and visual stimuli was set to each observer's preferred offset at the given orientation, as determined in Experiment 1. Surface depth was adjusted in 2 or 0.5 mm steps by pressing keys on the keypad (8 for deeper, 2 for shallower, and 5 to toggle between step sizes). Observers heard a long tone after each 2 mm adjustment and a short tone after each 0.5 mm adjustment. Observers could not set amplitude negative and were warned with a tone of any attempt to do so. The initial amplitude was set to a random value in the range 0-8mm (mean to peak). Three visual markers indicated the orientation along which to feel but the outer markers appeared without numbers. All other experimental details were as Experiment 1.

*3.2 Results and discussion*

There was no systematic variation in perceived surface amplitude with orientation for either plaid or single oblique stimuli (Figure 6). Importantly perceived depth amplitude did not approach zero (flat) for any participant at any orientation. With the exception of AT (min offset at 45°) and VC (did not participate in experiment 1), observers produced their smallest *inter-peak offsets* (see Fig 5) for stimuli oriented close to 0°, but there is no sign of a corresponding dip in *perceived depth amplitude* at 0° (or 45° for AT; Fig 6). To test for a systematic relationship between perceived depth amplitude and absolute inter-peak offset we correlated these judgments for the six participants who took part in both studies. A positive correlation would indicate that when observers aligned surface peaks with luminance peaks they also saw the stimulus as flat. With the exception of AS correlations were either very weak or negative and the mean correlation across the six observers was very weak and non significant ($\bar{r}$= 0.009 for the plaids and -0.03 for the single oblique stimuli). Thus we conclude that our inter-peak offsets are valid at all orientations and prediction 5 of Section 1.4 is rejected.

**Figure 6 about here (single column)**

## 4. Experiment 3: Perceived frequency is constant with stimulus orientation.

Experiment 1 had the limitation that observers could not adjust the frequency or shape of the haptic surface to match that of the visually perceived surface. The frequency of the haptic surface was always equal to that of the luminance signal and it was sinusoidal to match the luminance variations. This was done so as to reinforce a sinusoidal surface interpretation thus leaving the observers' internal illumination model as the only 'adjustable' parameter available to them in making their interpretations. Although there is evidence that humans readily perceive sinusoidal shading profiles as sinusoidal surfaces (Pentland, 1988; Schofield et al., 2006; Supplementary data) it is possible that our use of a haptic match stimulus forced observers into perceiving our stimuli in an unrealistic fashion. In particular, they may have wanted to report some stimuli as having a frequency half that of the shading as would be consistent with (say) vertical undulations lit from above (see Figure 1). In this experiment we asked observers simply to mark the locations of perceived surface peaks and troughs in the absence of any haptic cue to surface shape or frequency. Thus observers were free to perceive the surface as non-sinusoidal and as having a frequency different from that of the luminance signal.

### *4.1 Method*

Six participants from Experiment 2 (excluding VC) and four new observers (SH, TP,LA, & IH all with normal or corrected vision) viewed single, multiplicative (LM+AM), sinusoidal luminance modulations of the textured surface (see left hand side of Figure 3). Two variants of the experiment were conducted. The four new participants viewed stimuli in the haptic workstation although the Phantom device was not used and the stimuli were displayed and calibrated via a CRS-VSG2/5 graphics card. A modified hood which extended down to the edge of the mirror was used. Observers looked through a slit in this hood and as a result the viewing distance was reduced to 40cm. The seven observers from experiment 2 viewed stimuli outside of the haptic workstation on a vertically oriented 21" Sony GDM F520 monitor. The viewing distance was extended to maintain spatial frequency of the sinewave stimuli on the larger monitor. Otherwise, the experimental setup was identical to Experiment 1.

Sinusoidal shading profiles (LM+AM alone, sf=0.4 c/deg, see Figure 3) were presented at 12 orientations in the range 0-165°. Observers were instructed to mark the positions of peaks and troughs of the perceived surface by moving a red marker along a track defined by two blue markers (lower panel of Figure 4; markers shown as white and black respectively in print version). The position of the blue markers was chosen at random from trial to trial but their spacing was fixed (2.04 cycles of modulation) and the track was always orthogonal to the shading pattern. The red marker started 0.04 cycles away from one blue marker and

observers were told to mark features in order starting from this end of the range. Observers were, however, allowed to track back and forth to home-in on features. The position of the red marker was controlled using two bi-directional keys on a CRS CB3 button box (one each for coarse and fine adjustments). The third key was pressed up to mark a peak and down for a trough respectively. Thus, the direction of the marker key should have alternated on all trials.

The position of each marked location was recorded relative to the luminance profile of the visual stimulus. The distance between the marked features was also recorded. The data were screened to remove trials where the direction of the marker key did not alternate (e.g. where observers claimed to see two peaks without an intervening trough). The number of trials that were ignored due to this screening process was very small.

*4.2 Results and discussion*

The point- and diffuse-source assumptions lead to two predictions for illuminant aligned stimuli. A diffuse lighting interpretation would result in observers seeing a surface at the same frequency as the shading signal. Point-source model would promote frequency halving (undoing the frequency-doubling found for quadratic shading). Any perceptual flipping between these interpretations would alter the mean peak-to-trough spacing and increase standard deviations. The perceived distance between neighbouring peaks and troughs was close to ½ wavelength of the luminance signal at all orientations (Figure 7). Observers always perceived the surface undulations to have the same spatial frequency as the luminance signal regardless of stimulus orientation; there was no evidence for frequency halving at any orientation. There is no evidence that standard deviation varied systematically with orientation either suggesting that our observers saw a stable percept at all orientations. These results confirm those of Experiment 1 and allow us to reject the prediction that observers would perceive frequency halving at some orientations (prediction 1, Section 1.4).

Figure 8 shows the inter-peak offsets recorded for the four new observers in Experiment 3. Offset profiles are similar to those of Experiment 1 confirming that the previous result was unlikely to be due to the haptic matching method used. Comparing the results of the four new participants with those of the seven participants from Experiment 2 we see that neither past experience with the haptic task nor the exocentric inclination of the stimulus (backward slant of 45° in the haptic workstation) affect either the peak-to-trough spacing or the pattern of inter-peak offsets.

**Figures 7 and 8 about here (single column)**

## 5. Modelling.

We now propose two philosophically distinct models to explain our data. Noting that humans do not solve shape-from-shading veridically (Pentland, 1988; Zhang et al., 1999) we do not attempt to construct a machine vision algorithm to solve shape-from-shading to such precision. Many machine vision algorithms exist and interested readers are directed to Zhang et al. (1999) for an early review of such methods. It should be noted, however, that most of these methods assume a collimated (point-like) light source of known direction and require either iterative optimization of a cost function seeded with information such as occluding boundaries or the iterative propagation of information from seed points in the image such as intensity peaks. We avoided such methods because: (i) we are interested in human performance (not veridical shape recovery), (ii) our stimuli lacked many of the features that are required to make machine algorithms work, and (iii) we wanted to avoid methods that assume point-source lighting.

### 5.1 Model A: Assumed mixed illuminant.

Model A starts from the assumption that the human visual system has developed to process natural scenes and is thus optimised to a world that is mostly illuminated by a mixture of point and diffuse lighting or a least upwardly baised diffuse lighting (Dror et al., 2004; Mury, et al., 2009; Teller et a., 2001). The model assumes that human vision can, at least approximately, invert the generative processes that produce shading on a surface given some knowledge of the light source, and that when the stimulus provides few clues to the light source composition a default ilumination model is adopted in order that the inverse generative process can function. We draw a parallel here with Langer & Bülthoff (2000) who found that when a stimulus was rendered under point lighting observers mapped shape-from-shading as if under point lighting whereas surfaces lit diffusely were mapped according to a blur+dark-is-deep rule which is more appropriate for diffuse lighting.

In order to predict the default lighting adopted by each individual we generated the luminance profiles of physical sinusoidal surfaces lit by mixtures of point and diffuse lighting and then estimated the offset between the physical- and luminance-peaks by taking the Fourier transform of the luminance profile and extracting the phase of the component equal to the frequency of the original surface. This is equivalent to the blur imposed by Langer & Bülthoff. We then used MatLab's *fmincon* function, which finds the optimal constrained parameters for a arbitrary, user defined model with a user defined cost function (we used sum of squared errors as our cost function), to find the direction of the assumed point light source and balance between point and diffuse lighting that produced offset profiles that best

matched those for each observer (see Fig 5). These lighting models being fixed for all stimulus orientations.

The luminance profiles used to derive these fits were generated from Eqn 2 with two additional terms to describe the contribution from the diffuse source.

$$L \approx (1-\gamma)\left(\cos\left(\frac{\pi}{2}-e\right) - 0.12\sin(x)\cos\left(\frac{\pi}{2}-\varphi+\lambda\right)\sin\left(\frac{\pi}{2}-e\right) - \cos\left(\frac{\pi}{2}-e\right)\left(0.12^2 sin^2(x)\right)/2\right) + \gamma(0.065\cos(x) + 0.045\cos(2x))$$

Eqn5

Where $\gamma$ described the balance between point and diffuse lighting (high $\gamma$ = diffuse), $\varphi$ is the orientation of the surface corrugations (positive = clockwise), $\lambda$ is the direction of the default point source (positive = anti-clockwise) and the constants were appropriate to our stimuli (ie surface depth was 0.12 wavelengths and 0.065 and 0.045 provide the appropriate weighting for the first and second harmonics of the diffuse source modelled as decribed in the legend of Fig 2), $\omega$ is omitted as we assume it equal to 1. We further assumed elevation (e) = 30°.

The left hand side of Table 1 shows the SSE and parameter values for each observer. The model produces a relatively good fits to the data although data from those observers having a smooth offset profile with a high peak were not fit well. Resulting fits are shown by the solid lines in Figs 5 and 8.

Table 1 about here

5.2 Model B: Mixed processing model.

Model A relies on the observer being able to at least approximately invert the generative process in order to estimate shape-from-shading; it does not articulate a means by which this is achieved. Given that shape-from-shading estimates are often not veridical this inversion seems unlikely. We now present an alternative, mechanism driven, account of our data.

5.2.1 Outline

Model B starts with the assumption that humans process all stimuli with two shape-from-shading modules whose output is then combined in a stimulus specific way. This combination could be the result of flipping between two hypothesised surface shapes but, given our data, we think a linear combination of the two hypothesised surface is more likely.

We implement a cut-down version of Pentland's (1988) model for human shape-from-shading which produces a linear mapping between luminance and perceived surface shape with a ¼ wavelength offset. We augment this model with a version of Langer & Bultholf's (2000) blur+dark-is-deep model. These two models give a reasonable account of human shape-from-shading under point- and diffuse-lighting assumptions respectively. Critically when presented with sinusoidal shading patterns they will both produce sinusoidal depth profiles but Pentland's (1988) model will shift the perceived surface peak by ¼ wavelength relative to the luminance peak[3] whereas the two peaks will align in the output of Langer and Bulthoff's (2000) model. The principal innovation of this model is to combine the two approaches above such that, when fit to the data, the balance between the point- and diffuse-lighting interpretations can be inferred. The use of sinusoidal stimuli greatly simplifies the model. Because each sub-module will produce a sine wave output we need not implement the models in full but can simulate their action with appropriately phase shifted sine waves.

**Figure 9 about here (double column)**

The upper arm of Model B also provides an estimate of the lighting direction..Note that although sinusoidal shading is highly ambiguous with relation to the direction of the light source Koenderink et al. (2007) have shown that anisotropic shading patterns give rise to very stable estimates of illumination direction up to 180° flips. In this case people estimate the light direction to be orthogonal to the dominant orientation in the shading pattern. In practice there are two directions orthogonal to the dominant orientation in each stimulus; both are equally valid estimates and we deal with this ambiguity in section 5.2.2.
The two surface interpretations are combined in a weighted sum. Each arm has a weight ($\beta$ and 1-$\beta$ for the diffuse and point interpretations respectively) and $\beta$ is fixed for each observer. The point interpretation has an additional variable weight which depends on the observer's estimate of the likelihood that illumination will come from the direction specified by the stimulus. If an observer had a preference for lighting from above (say) this would be expressed as a strong weight for vertical lighting and a weak weight for horizontal lighting.

When presented with a sinusoidal input the two arms of this model will produce sinusoidal surface profiles: one offset by ¼ wavelength from the luminance profile, the other having no offset. The weighted sum of two sine waves with the same frequency but different phases is

[3] A model based on integration (surface gradients proportional to luminance) would also produce a ¼ wavelength offset.

a sine wave with intermediate phase. For the purposed of our fits the model is described by the following simple equation $s = \cos(x + \arctan(\cos(\theta - \lambda) \cdot (1 - \beta) / \beta))$, where $\theta$ is orthogonal to the dominant orientation in the stimulus - and $\lambda$ is the preferred lighting direction which, unlike $\theta$, indexes anticlockwise. Thus the inter-peak offset predicted by the model will be closer to whichever of the two interpretations carries the stronger weight. As the relative weighting of the two components varies with stimulus orientation so does the inter-peak offset.

*5.2.2 Direction-dependent weighting function*

We made the variable weighting function sinusoidal such that a negative weight would be assigned if the illumination was predicted to come from the direction opposite to the observer's preferred direction. Recall that there are two directions orthogonal to a given stimulus orientation and both are candidates for the perceived lighting direction. In our framework lighting from one direction would produce a positive inter-peak offset relative to its own direction whereas lighting from the opposite direction would produce a negative inter-peak offset relative to its own direction – lighting direction repeats every 360°. However, in terms of the stimulus both the predicted offsets will be in the same direction because stimulus orientation repeats every 180°. The weighting function can thus produce negative weights and hence negative offsets allowing us to model the data of participants like AT. The orientation with the most positive weight is deemed to be the observer's preferred lighting direction and this is adjusted to fit the data best.

*5.2.3 Analysis and Results*

We fit (using *fmincon*) the model to the individual data from Experiments 1 & 3 dashed-lines in Figures 5 and 8). Note that the dashed lines exactly overlay the predictions from Model A. Model parameters and SSE's for all observers are shown on the right hand side of Table 1. Note that the two models produce nearly identical λ's and SSE value; although β and $\gamma$ are not identical they are perfectly correlated. These results strongly suggest that the two models are mathematically equivalent, a fact we prove in Supplementary section S3.

Model B (see Fig 9) implies that perceived depth amplitude will vary with orientation because the amplitude of *s* depends on *w*. We did not observe any such variation which is a challenge to the specific form of the model although this anomally can be reconciled by supposing that the amplitude of surface *s* is normalised to the stimulus contrast.

5.3 Interpretation

762 The parameters for Model A should be interpreted as follows. The direction of the observers
763 preferred light source (but not its elevation) is given by λ (Note that lighting direction is
764 indexed anti-clockwise whereas stimulus orientation indexes clockwise) and the balance
765 between the diffuse and point components is deterimined by $\gamma$ such that high $\gamma$ suggests a
766 mostly diffuse default illuminant. For Model B *λ* again the observers preferred light source
767 and *β* is the weighting term. A high $\gamma$ *(β*) means that the observer prefers a diffuse source
768 interpretation. A low $\gamma$ *(β)* implies that a point source is preferred when viable.
769
770 Low $\gamma$ or *β* also results in a flat-topped model offset profile with an abrupt transition between
771 extreme offsets (eg. observer PDJ in Figure 5). High $\gamma$ or *β* results in smoother transitions
772 and a lower maximum offset. Note that where an observer's maximum inter-peak offset is
773 large (close to ¼ wavelength) the models will tend to prefer a low $\gamma$ or *β*, resulting in abrupt
774 transitions. Thus observers with a large maximum offset but smooth transitions present a
775 challenge to the models. For Model B at least data from these observers might be better fit
776 by assuming a weighting function of a different shape. There was considerable variation in $\gamma$
777 and *β* across participants, suggesting that some preferred the diffuse lighting interpretation
778 more than others.
779
780 *5.4 Dealing with plaid stimuli*
781 The above models consider only stimuli comprising single sine wave luminance profiles. The
782 stimuli used in Experiment 1 were more complex plaid stimuli in which one orientation
783 faithfully represented multiplicative shading (LM+AM) while the other did not (LM-AM). We
784 have shown elsewhere that observers treat LM-AM as if it were a flat reflectance change
785 (Schofield et al, 2006; Schofield et al., 2010). Layer segmentation – the separation of
786 components into shading vs reflectance changes – is a complex issue in itself but humans
787 seem to be able to perform such a separation (Kingdom, 2008). We assume that layer
788 segmentation takes place before shape-from-shading such that our plaid stimuli present
789 themselves as single sine waves as far as shape-from-shading is concerned. Elsewhere we
790 propose a model for how layer segmentation is achieved in LM/AM plaids (Schofield, et al.,
791 2010). Layer segmentation will also separate the noise textures in our stimuli from the
792 shading patterns.
793
794 *5.5 Convex/concave ambiguities and high $\gamma$ (β)*
795 In the introduction we outlined five predictions. One (prediction 3) concerned the
796 convex/concave ambiguity and the possibility of perceptual flipping between two equally
797 likely surface profiles. We noted that this would predict small inter-peak offsets for some
798 stimuli but that it was also make the data these orientations bimodal. We discounted this

hypothesis in section 2.3 because we found a positive relationship between absolute offsets and bimodality. That is, bimodality was associated with large offsets not small offsets. However, the analysis of Experiment 1 merged the results from observers such as PDJ with abrupt transitions and those of observers such as HS with smooth transitions. Is it possible that only those with smooth transitions suffer perceptual flipping and that this explains the smoothness of their offset data? We reasoned that if people with smooth transitions (high $\gamma$) suffered perceptual flipping more than those with abrupt transitions (low $\gamma$) then the individual offset-bimodality correlations measured in Experiments 1 and 3 should themselves correlate negatively with $\gamma$. Recall that offset-bimodality correlations will be negative if perceptual flipping occurs at orientations with small offsets. Although this relationship was negative it was relatively weak and not statistically significant ($r$=-26, $df$=17, $p$=0.283 ). We also tested the correlation between people's mean coefficient of bimodality and $\gamma$ which should be positive if flipping/bimodality is the cause of smooth offset profiles (high $\gamma$). This correlation was very weak negative and not significant ($r$=-0051, $df$=17, $p$=.84). Finally we measured the correlation between $\gamma$ and coefficients of bimodality associated with individual's smallest offsets. Again this should be positive if perceptual flipping is causing smooth offset profiles; it was not ($r$=-.11, $df$=17, $p$=.64). We conclude, as in section 2.3, that that the concave/convex ambiguity was not responsible for producing the smooth offset profiles and high $\gamma$ values noted in our data.

## 6. Discussion.

People perceive sinusoidal luminance shading as a sinusoidal surface undulating at the same spatial frequency as the luminance profile (see Schofield, at al., 2006, Pentland 1988, Experiment 3, and supplementary file); dismissing prediction (1), see section 1.4.

Perceived inter-peak offsets varied systematically with orientation. This result is not consistent with the assumption of a single, pure point source (even one with variable direction; prediction 2), since that would predict no change in inter-peak offset with stimulus orientation. This finding is not consistent with a fully diffuse light source either, since that would predict no offset at any orientation. The variation in inter-peak offsets was not accompanied by a reduction in perceived depth amplitude, nor was it due to perceptual flipping between multiple, equally likely, surface interpretations; so predictions 3 and 5 are also dismissed.

Our data can be modelled by assuming the observer is able to, at least approximately, reverse the image generation process using a mixed, but fixed, internal lighting assumption (Model A) or that they generate two surface interpretations which are linearly combined with

836 weights determined by the stimulus (Model B). In either case model fits suggest that
837 observers adopt a mixture of point and diffuse lighting. The two models are mathematically
838 equivalent for sine wave shading patterns so our data cannot discriminate between them.
839
840 Our finding that observers seem to adopt a mixed point and diffuse lighting model is
841 consistent with the results of lightness judgements found by Bloj, Ripamonti, Mitha, Hauck,
842 Greenwald & Brainard (2004). A mixed lighting model is also consistent with the data on
843 natural illumination which show a largely diffuse illumination with an upward bias – that is a
844 combination of diffuse and directional components (Dror, et al., 2004; Mury et al, 2009;
845 Teller, et al., 2001). It would make sense if humans adopted an illumination model which
846 was close to the naturally occurring illumination profile. The inclusion of an explicit point light
847 source (rather than a more general upward bias as might be more common in natural
848 settings; Mury et al, 2009) facilitated matches to individual data. It seems likely that
849 individuals have a preferred lighting direction that is generally from above but which varies
850 between observers and can be modified by experience (Adams et al., 2004; Mamassian &
851 Goutcher, 2001; Sun & Perona, 1998). This suggests to us that a discrete point component
852 rather than a general upward bias is appropriate.
853
854 Model B is limited to the understanding of sinusoidal shading. It could be expanded to deal
855 with (that is, ignore) reflectance changes by the inclusion of a preceding layer-segmentation
856 stage (see for example that proposed by Schofield et al., 2010). It might also be extended to
857 more complex natural patterns by implementing the linear shading (Pentland, 1988) and
858 blurred dark-is-deep (Langer and Bülthoff, 2000) models in full. A method based on
859 Pentland's (1982) model for finding the illumination direction could serve to expand the
860 illuminant direction estimation process to more natural images (see also Gerhard & Maloney,
861 2010). It would be interesting (but beyond the scope of the current paper) to test such a
862 model against human performance for more complex scenes. Pentland's (1988) model alone
863 does reasonably well in such situations.
864
865 Model A is similarly limited to surfaces with uniform albedo and can also be augments by a
866 preceding layer-segmentation stage. In theory this model can deal with any type of surface
867 however in practice any implementation would require that the image generation process be
868 inverted. This amounts to solving the shape-from-shading problem given an assumed light
869 source which may prove pragmatically difficult for the general case.
870
871 Our use of sinusoidal shading patterns may over-emphasise the diffuse lighting component.
872 Our stimuli contain no sharp edges that might indicate hard shadows and thus the presence

873 of a point light source. Further, our stimuli may promote the perception of a Lambertian
874 surface with little or no specular component. Images with more obvious specular highlights
875 may require a different interpretation from the one outlined here. However, our models are
876 more generally applicable if we allow the $\gamma$ (or $\beta$) to vary with stimulus content such as hard
877 edges and specular highlights. Schemes such as Freeman's (1994) generic view framework
878 might serve to adjust $\gamma$ in more complex scenes if diffuse lighting were included as a
879 candidate lighting model. Non-sinusoidal shading profiles, especially those with occlusions,
880 might indicate harsher – more point-like – lighting, giving the point source component of the
881 model a greater weight. We note that people are rather good at estimating the true light field
882 in well articulated, object rich, scenes (Koenderink et al., 2007) and that in such cases
883 internal lighting biases may not apply at all. However, while some stimulus types might
884 provide little evidence that there is a diffuse component within the human default lighting
885 model, we argue that the most general model must contain such a component.
886
887 A potential criticism of our method is that people may not actually perceive our stimuli as
888 conveying realistic depth percepts but might rather associate luminance variations with
889 depth in order to follow the instructions given; an experimenter effect. We reject this for three
890 reasons. First, it is unlikely that all of the naive participants would adopt the same 'false'
891 association between luminance and depth to please the experimenter and that *none* would
892 set their depth / gradient estimates to zero if they in fact saw our stimuli as flat. Second, if
893 observers had adopted a simple association between luminance and depth we think it
894 unlikely that this association would vary systematically with stimulus orientation. Third, we
895 have shown elsewhere (Schofield et al., 2010) that observers see the LM-AM components of
896 our plaid stimuli as flat. The LM-AM and LM+AM components contain the same luminance
897 variation and the AM sub-components create relatively subtle variations in pixel-wise
898 luminance values. Therefore it seems unlikely that a 'false' association between luminance
899 and depth would be applied to LM+AM stimuli alone and much more likely that observers
900 genuinely perceive LM+AM as conveying depth.
901
902 7. Conclusion
903 People perceive sinusoidally corrugated luminance patterns as sinusoidal surfaces of the
904 same spatial frequency as the luminance waveform. In general perceived surface peaks are
905 offset from the luminance peaks and these inter-peak offsets vary with stimulus orientation.
906 This result is not consistent with an internal lighting model that is either a pure point source
907 or fully diffuse illumination, but it is consistent with a weighted mixture of the two lighting
908 types. Such as mixed illumination model is consistent with everyday experience of biased

diffuse illumination as found on a cloudy day, in the illumination field of a typical room with an overhead light and light-coloured walls, or from the sun embedded in a diffusing sky.

**Acknowledgments**

This work was funded by EPSRC grants GR/S07254/01 & EP/F026269/1 to AJS & GR/S07261/01 to MAG. EPSRC took no part in the study or manuscript preparation. We thank Harriet Allen and Martin Banks for commenting on early drafts of the manuscript, and Mike Landy and an anonymous reviewer for their helpful suggestions.

*Author contributions.* AJS devised the study, collected some of the data for experiments (1 &2), devised and implemented experiment 3 and the control experiment, conducted much of the analysis and was the principle author of the manuscript. PBR designed and implemented experiments 1 and 2 and collected data for them. MAG advised on study design and made considerable revisions to the final manuscript.

**References**

Adams, W.J., Graf, E.W., & Ernst, M.O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7*, 1057-1058.

Belhumeur, P.N., Kriegman, D.J., Yuille, A.L., (1999) The Bas-Relief Ambiguity, *International Journal of Computer Vision, 35,* 33-44.

Bloj, M., Ripamonti, C., Mitha, K., Hauck, R., Greenwald, S., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal of Vision, 4*(9):6, 735-746, http://journalofvision.org/4/9/6/, doi:10.1167/4.9.6.

Brainard, D. H., Pelli, D.G., & Robson, T. (2002). Display characterization. In *Encyclopedia of Imaging Science and Technology*. J. Hornak (ed.), pp.172-188, Wiley.

Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios, and intaglios into cameos, with and account of other analogous phenomena. *Edinburgh Journal of Science, 4*, 99-108.

Cavanagh, P., & Leclerc, Y.G. (1989) Shape from shadows. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 3-27.

945 Christou, C.G., & Koenderink, J.J. (1997). Light source dependence in shape from shading.
946 *Vision Research, 37*, 1441-1449.
947
948 Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004). Statistical characterization of real-world
949 illumination. *Journal of Vision, 4*(9):11, 821-837, http://journalofvision.org/4/9/11/,
950 doi:10.1167/4.9.11.
951
952 D'Zmura, M., (1991) Shading ambiguity: Reflectance and Illumination, in *Computational*
953 *models of visual processing,* M.S. Landy, & J.A. Movshon (eds), pp 187-207, MIT Press,
954 Cambridge, MA.
955
956 Erens, R.G.F., Kappers, A.K.L., & Koenderink, J.J. (1993). Perception of local shape from
957 shading. *Perception & Psychophysics, 54*, 145-156.
958
959 Freeman, W.T. (1994). The generic viewpoint assumption in a framework for visual
960 perception. *Nature, 368*, 542-545.
961
962 Gerhard, H.E., Maloney, L.T. (2010). Estimating changes in lighting direction in binocularly
963 viewed three-dimentional scenes. *Journal of Vision*, 10(9):14,
964 http://www.journalofvision.org/content/10/9/14
965
966 Hess, E.H. (1950). Development of chicks' responses to light and shade cues to depth.
967 *Journal of Comparative and Physiological Psychology, 43*, 112-122.
968
969 Kingdom, F.A.A. (2003) Colour brings relief to human vision. *Nature Neuroscience, 6*, 641-
970 644.
971
972 Kingdom, F.A.A. (2008) Perceiving light versus material. *Vision Research 48*, 2090-2105.
973
974 Kleffner, D.A. & Ramachandran, V.S. (1992). On the perception of shape from shading.
975 *Perception & Psychophysics, 52*, 18-36.
976
977 Koenderink, J.J., & Pont, S.C. (2003). Irradiation direction from texture. *Journal of the*
978 *Optical Society of America A – Optics Image Science and Vision, 20*, 1875-1882.
979
980 Koenderink, J.J., Pont, S.C., van Doorn, A.J., Kappers, A.M.L., & Todd, J.T. (2007). The
981 visual light field. *Perception, 36,* 1595-1610.

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

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

Koenderink, J.J., van Doorn A.J., & Pont, S.C. (2007). Perception of illuminance flow in the case of anisotropic rough surfaces. *Perception & Psychophysics, 69*, 895-903.

Langer, M.S., & Bülthoff, H.H. (2000). Depth discrimination from shading under diffuse lighting. *Perception, 29*, 649-660.

Langer M.S. & Zucker, S.W. (1997). Shape-from-shading on a cloudy day. *Journal of the Optical Society of America A – Optics Image Science and Vision, 11*, 467-478.

Liu, B., & Todd, J.T., (2004). Perceptual biases in the interpretation of 3D shape from shading. *Vision Research, 44*, 2135-2145.

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

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

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*, 565-570.

Pentland, A. (1988) Shape Information From Shading: A Theory About Human Perception. *Second International Conference on Computer Vision*, 404-413.

Pentland, A.P. (1982) Finding the illuminant direction. *Journal of the Optical Society of America*, 72, 448-455.

Ramachandran, V.S. (1988). Perception of shape-from-shading. *Nature, 331*, 163-165.

Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society, 2*, 37-42.

Schofield, A.J., Hesse, G., Rock P.B., & Georgeson, M.A. (2006). Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces. *Vision Research, 46*, 3462-3482.

Schofield, A.J., Rock, P.B., Sun, P., Jiang, X., & Georgeson. (2010). What is second-order vision for? Discriminating illumination versus material changes. *Journal of Vision, 10*(9): 2, http://www.journalofvision.org/content/10/9/2, doi: 10.1167/10.9.2.

Stewart, A.J., & Langer, M.S. (1997). Towards accurate recovery of shape from shading under diffuse lighting*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*, 1020-1025.

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience,1*, 183-184.

Teller, S., Antone, M., Bosse, M., Coorge, S., Jethwa, M., & Masters, N. (2001). Calibrated, registered images of an extended urban area. *Proceedings of the IEEE computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii.

Todd, J.T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and performance, 9*, 583-595.

Tyler, C.W. (1998). Diffuse illumination as a default assumption for shape-from-shading in graded images. *Journal of Image Science and Technology, 42*, 319-325.

von Fieandt, K. (1949). The phenomenological problem of light and shadow. *Acta Psychologica, 6*, 337-357.

Wijntjes, M.W.A., Volcic, R., Pont, S.C., Koenderink, J.J., & Kappers, A.M.L. (2009). Haptic perception disambiguates visual perception of 3D shape*. Experimental Brain Research, 193*, 639-644.

Zhang, R., Tsai, P-S., Cryer, J.E., Shah, M. (1999) Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*, 690-706.

Figure and Table Legends

Figure 1. A: Luminance profiles for sinusoidal surfaces under a point-source illuminant. Here and throughout the paper we take point-source to mean a highly concentrated but distant light source. Outer ring: example rendered stimuli. Inner ring: surface profiles (thin lines) and shading (luminance) profiles (thick lines) for each rendering. Sub-plot the x-axes represent position along the surface; y-axes represent luminance (thick lines) or height (thin lines). Sub-plots trace surface depth and luminance from left to right working along a lines from A to B in the images of the outer ring. This configuration is counter intuitive for some plot pairs but maintains a common reference frame. Note that as surface orientation repeats every 180°, plots on the left of the figure mirror those on the right. Values of $\psi$ indicate the offset between physical- and luminance peaks in wavelengths. The polar location of each inner plot represents the orientation of the surface (polar angle, see panel C) and elevation of the light source relative to the centre of the surface (radial distance from centre) where frontal lighting would be represented by a plot at the centre of the figure and oblique lighting by a plot on the outer circle. For example, the top most image and associated plot represent a vertical surface lit from above, the images and plots at 90 and 270° represent horizontal stimuli also lit from above. The inner plots are located at a distance from the centre of the figure appropriate for the elevation of the light source used in each rendering. These were 45° for surfaces oriented at 90 and 270°, 41° for surfaces at 60,120,240 and 300°, 27° for surfaces at 30,150, 210 and 330°, and 30° for surfaces at 0 and 180°. With the exception of 0 and 180° surfaces, light elevations were chosen to avoid occlusions and double-crested peaks (see supplementary file). The depth amplitude (mean-to-peak) of the surfaces was 0.12 of the undulation wavelength, matching that used in experiment 1. Rendered images are from *PoVRay* (Persistence of Vision Raytracer Pty. Ltd), and traces from *MatLab*, both assuming Lambertian shading. B: Lighting diagram depicting variable lighting direction d. C: Lighting diagram showing variable orientation for the physical surface.

Figure 2. Sinusoidal surface under diffuse illumination. a) surface as described in Figure 1 rendered (using *PovRay*) under a spherical diffuse illumination model consisting of a spherical array of 400, randomly but evenly spaced light sources. Minor fluctuations in gray level are due to the sampling process. b) surface (dashed line, left axis) and luminance (solid line, right axis) traces for the central 2 cycles of a similar surface rendered in MatLab under a diffuse source sampled at 1568 random positions in front of the surface. The strength of each light in the latter diffuse model was 1/1568th of that for the source of Figure 1.

Figure 3. As figure 1 except rendered images are sub-plot luminance traces now show the case of mixed diffuse and point source lighting (weighted 0.75 diffuse, 0.25 point).

Figure 4. Extracts from example stimuli arranged as Figure 1. To save space only the cardinal and 45° oblique orientations are shown. Stimuli labelled with dashed lines (top and right hand column) are from Experiments 1 and 2 where plaid stimuli were used those with solid lines (bottom and left hand column) from Experiments 2 and 3 where single gratings were used. Radial lines have been labelled to reflect stimulus orientation in the range 0-180°. These stimuli have been cropped for presentation, un-cropped versions are shown in supplementary Figure S3. No gamma correction has been applied to these stimuli but the noise contrast has been exaggerated to aid visualisation. However, despite this manipulation, we note that the example stimuli do not provide an especially good representation of the appearance of our stimuli within the lab setting. Specifically we are aware that people find it harder to perceive depth in our stimuli when presented in paper form than is the case during experiments. See Section 4.1 for a description of the markers on bottom panel (coloured white and blank in print but red and blue in the experiment and in the online version). The stimuli are best viewed online at 200% magnification.

Figure 5. Example inter-peak offset data from nine observers as a function of stimulus orientation (measured clockwise from vertical). Circles represent recorded data; triangles are extrapolated data (see Analysis). Lines represent model fits (see section 5). Error bars are standard deviations. AJS was an author.

Figure 6. Perceived surface amplitude measured as the distance between the zero-crossings of the haptic sine wave (dc position of the surface) and the haptic surface peaks in the direction normal to the surface plane. Traces show results for individual observers as a function of orientation. AS was not an author.

Figure 7. Peak-to-trough spacing versus stimulus orientation. Data points show the mean spacing between neighbouring peaks and troughs in the perceived surface at different orientations averaged across repeated trials. Data for observers TP to HS have been shifted vertically in integer steps for clarity. Error bars represent 95% confidence intervals.

Figure 8. Inter-peak offset data from Experiment 3. Details as Figure 5.

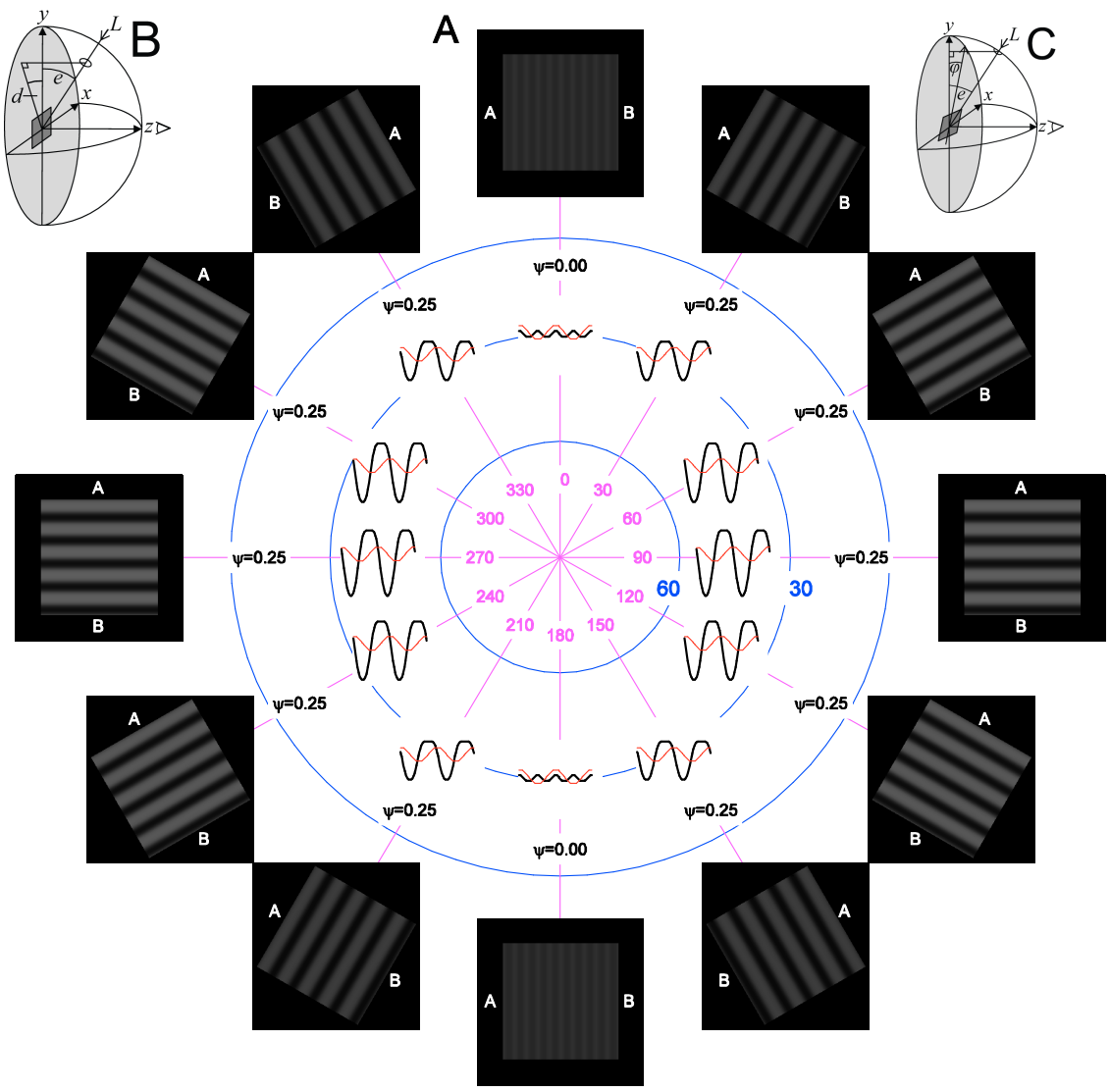
Figure 9. A) Lighting diagram showing orientation of stimulus φ and orientation of default point light source λ for Model A. B. Schematic diagram of Model B simplified for the case of sine-wave stimuli. Estimates of surface shape are predicted for a point source interpretation

(upper arm) and a diffuse source interpretation (lower arm). These are combined with a variable weight determined by an estimate of the lighting direction.

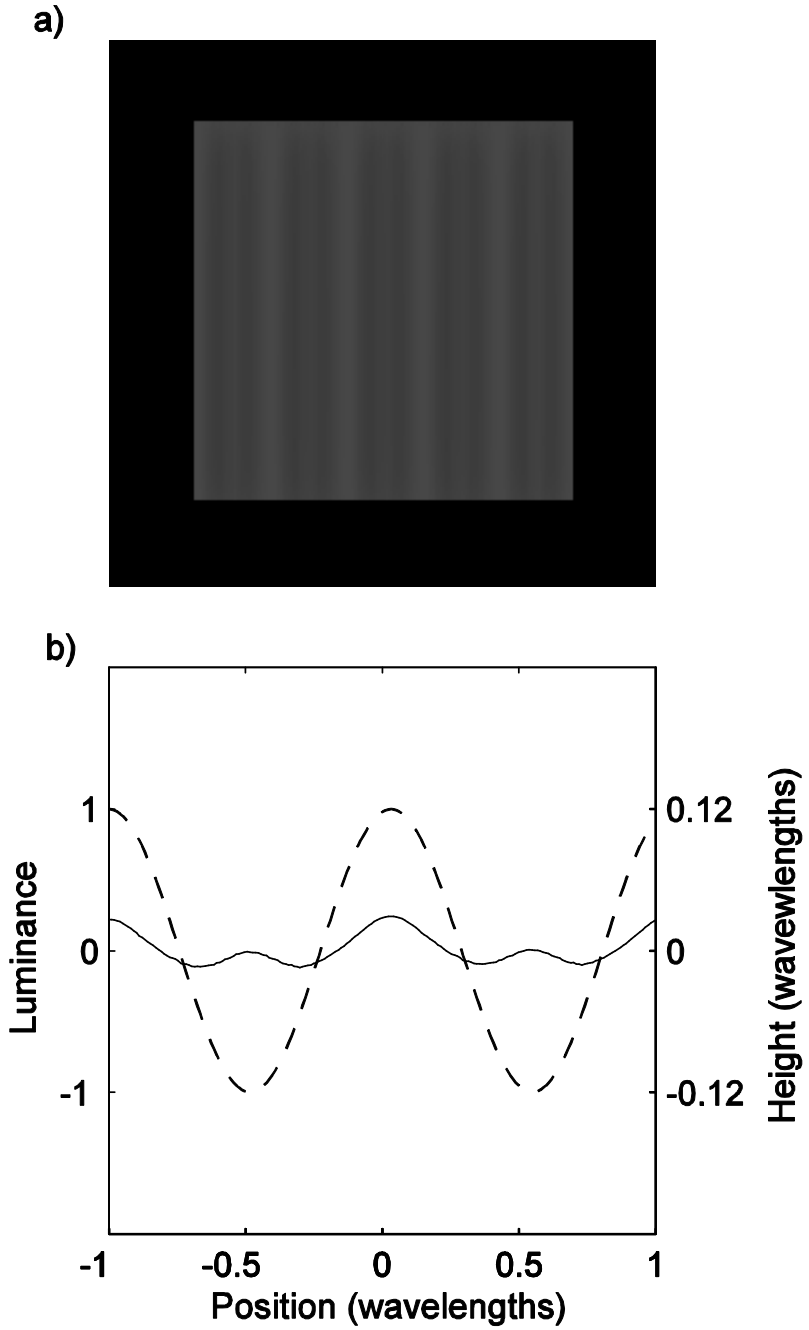
| | Model A | | | Model B | | |
|---|---|---|---|---|---|---|
| Person | $SSE_A$ | γ | Preferred light source ($\lambda_A$) | $SSE_B$ | β | Preferred light source ($\lambda_B$) |
| AJS | 0.005 | 0.45 | 6 | 0.005 | 0.34 | 6 |
| PDJ | 0.048 | 0.21 | -25 | 0.048 | 0.14 | -25 |
| RCL | 0.013 | 0.41 | 5 | 0.013 | 0.3 | 5 |
| AO | 0.033 | 0.34 | 6 | 0.033 | 0.25 | 6 |
| HW | 0.071 | 0.23 | 5 | 0.071 | 0.16 | 5 |
| AS | 0.081 | 0.64 | -12 | 0.081 | 0.53 | -12 |
| HS | 0.005 | 0.54 | 3 | 0.005 | 0.42 | 3 |
| AT | 0.061 | 0.33 | 132 | 0.061 | 0.23 | 132 |
| KU | 0.056 | 0.78 | 151 | 0.053 | 0.67 | 129 |
| AC | 0.004 | 0.69 | 31 | 0.004 | 0.59 | 31 |
| JG | 0.019 | 0.38 | -13 | 0.019 | 0.28 | -13 |
| PS | 0.027 | 0.5 | 6 | 0.027 | 0.38 | 6 |
| PR | 0.006 | 0.43 | -4 | 0.006 | 0.32 | -4 |
| MH | 0.049 | 0.36 | -153 | 0.049 | 0.26 | -153 |
| SW | 0.056 | 0.64 | -46 | 0.056 | 0.52 | -46 |
| SH | 0.012 | 0.09 | 13 | 0.012 | 0.06 | 13 |
| TP | 0.019 | 0.53 | -5 | 0.019 | 0.41 | -5 |
| LA | 0.009 | 0.54 | -3 | 0.009 | 0.43 | -3 |
| IH | 0.015 | 0.28 | -7 | 0.015 | 0.2 | -7 |

Table 1.

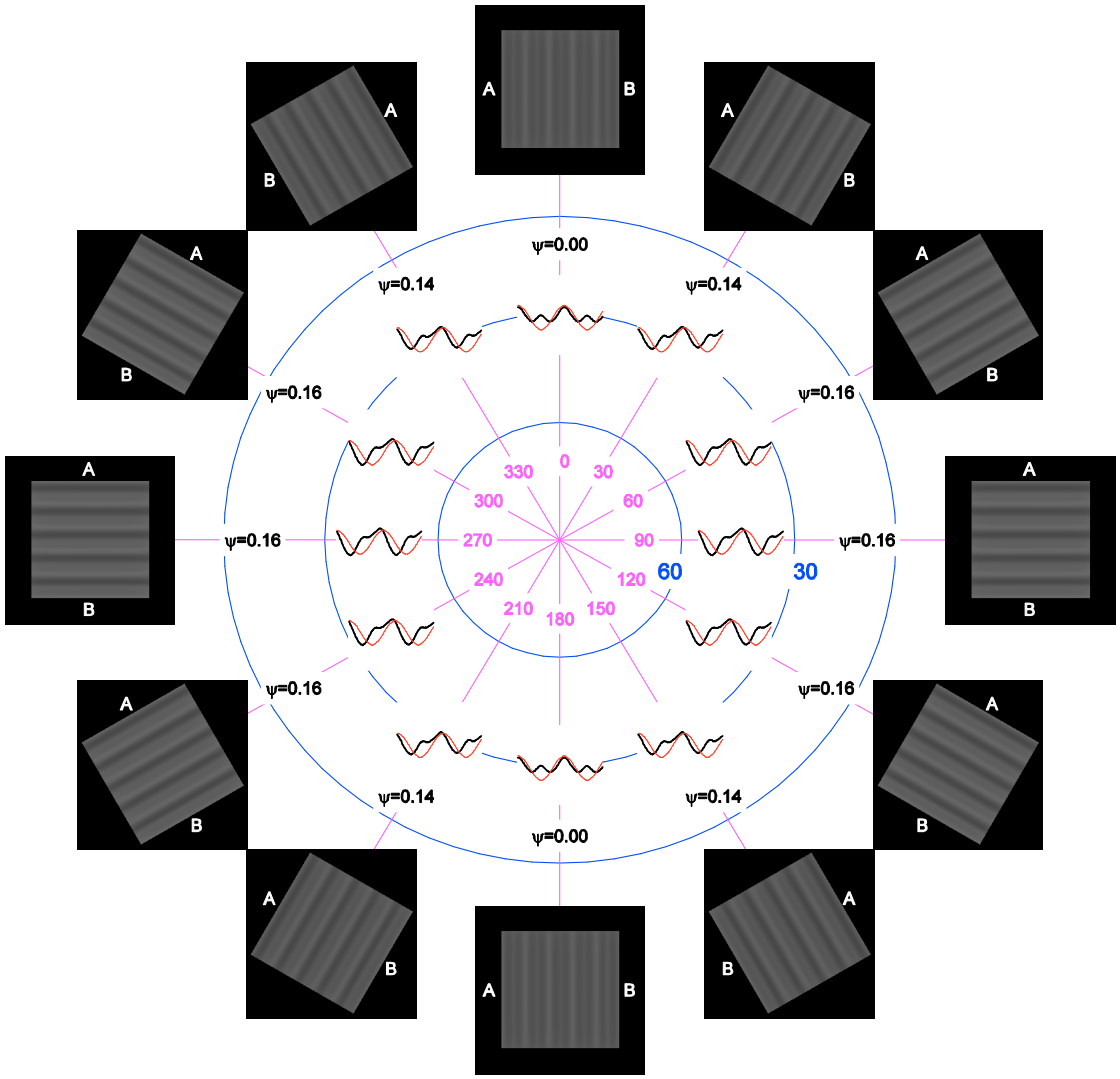
Model fit parameters. Model A, left-hand side: columns show the sum of squared errors between the modelled and observed inter-peak offsets, the weight (γ) applied to the diffuse interpretation, and the observers' preferred lighting direction (λ). Model B, right-hand side: SSE, weight $\beta$ and $\lambda$ . Lighting direction is given as positive = anti-clockwise shift from vertical. AJS and PR are authors.



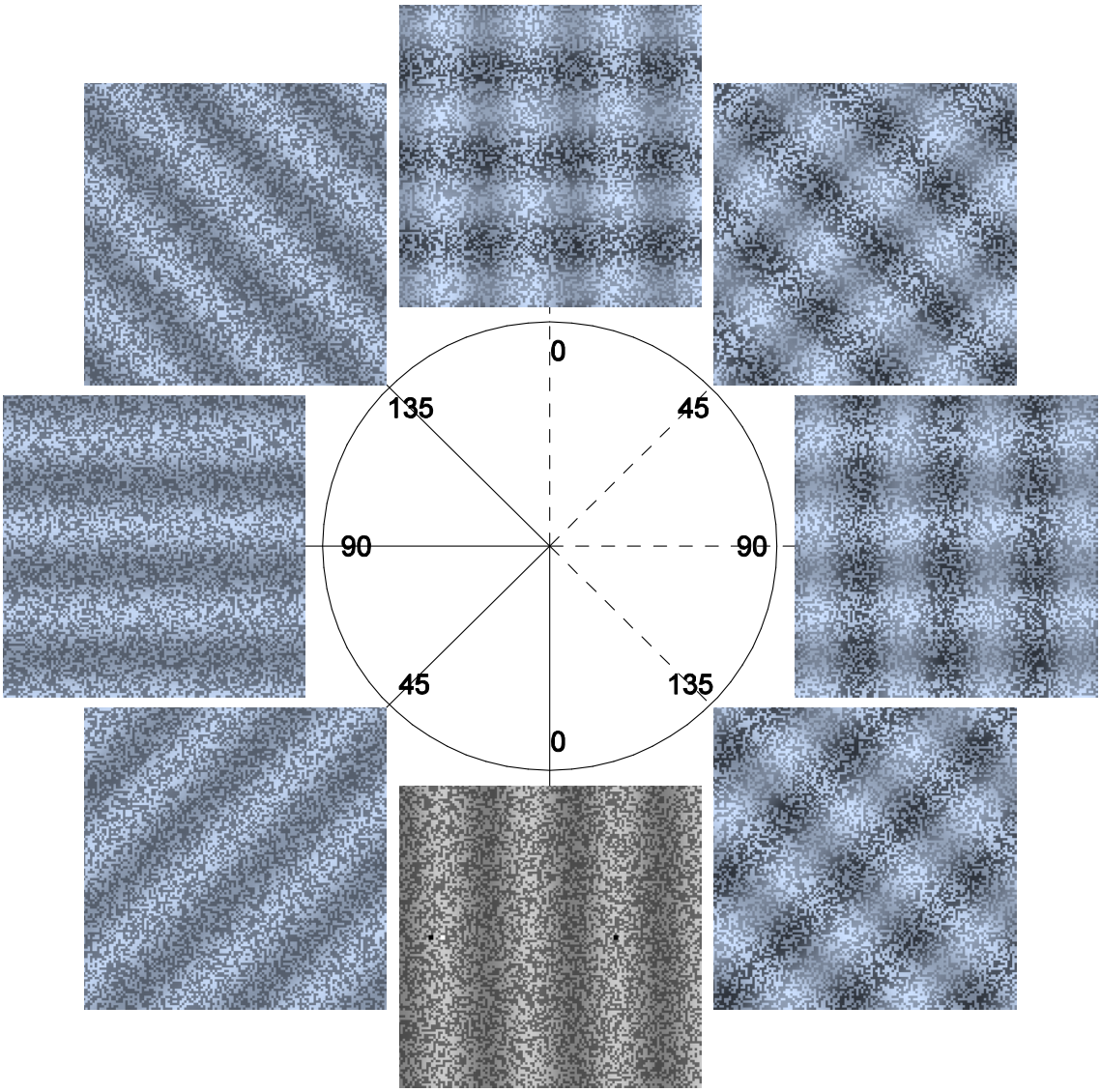

1152
1153 Fig 1



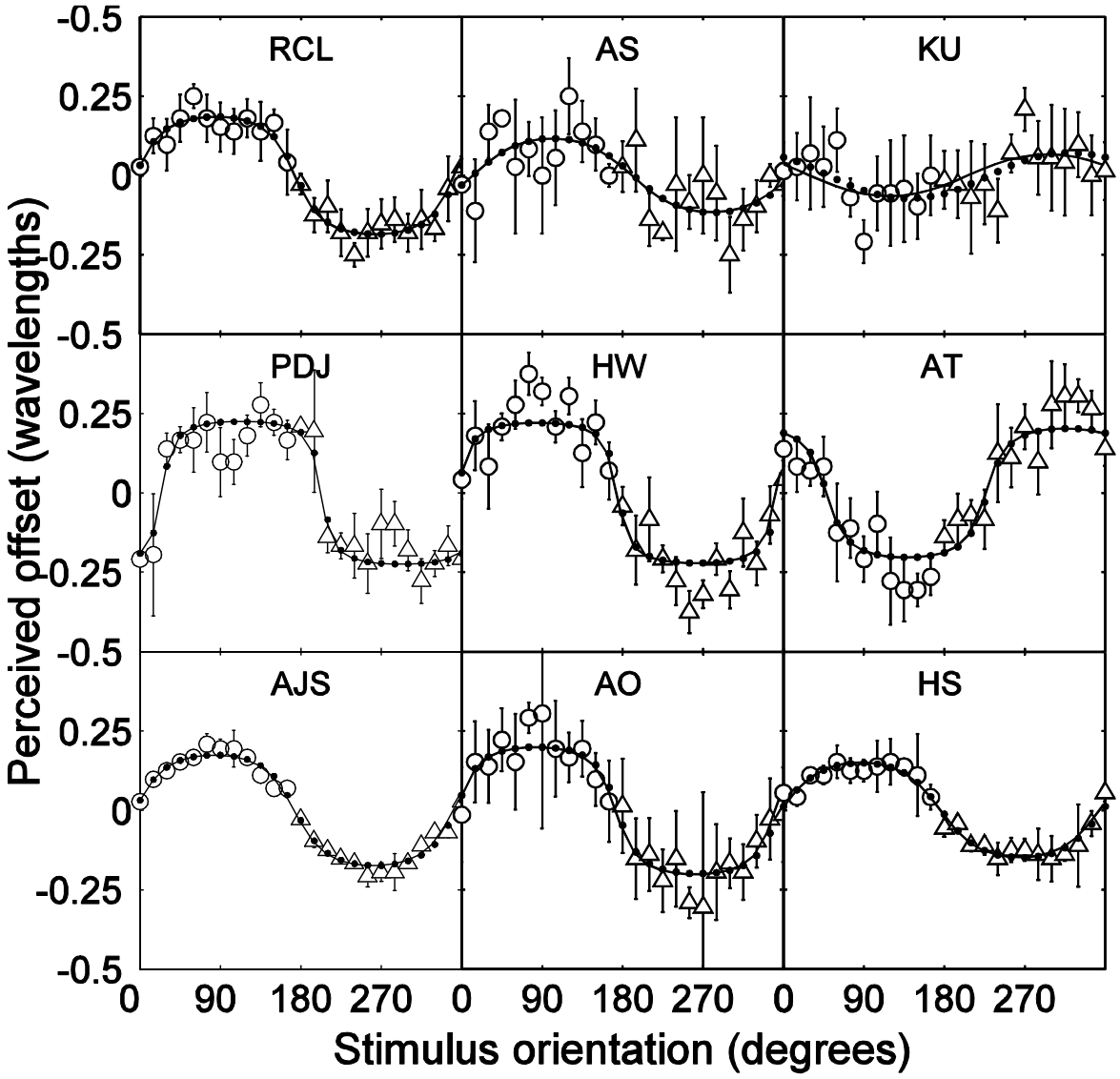

1154
1155 Fig 2



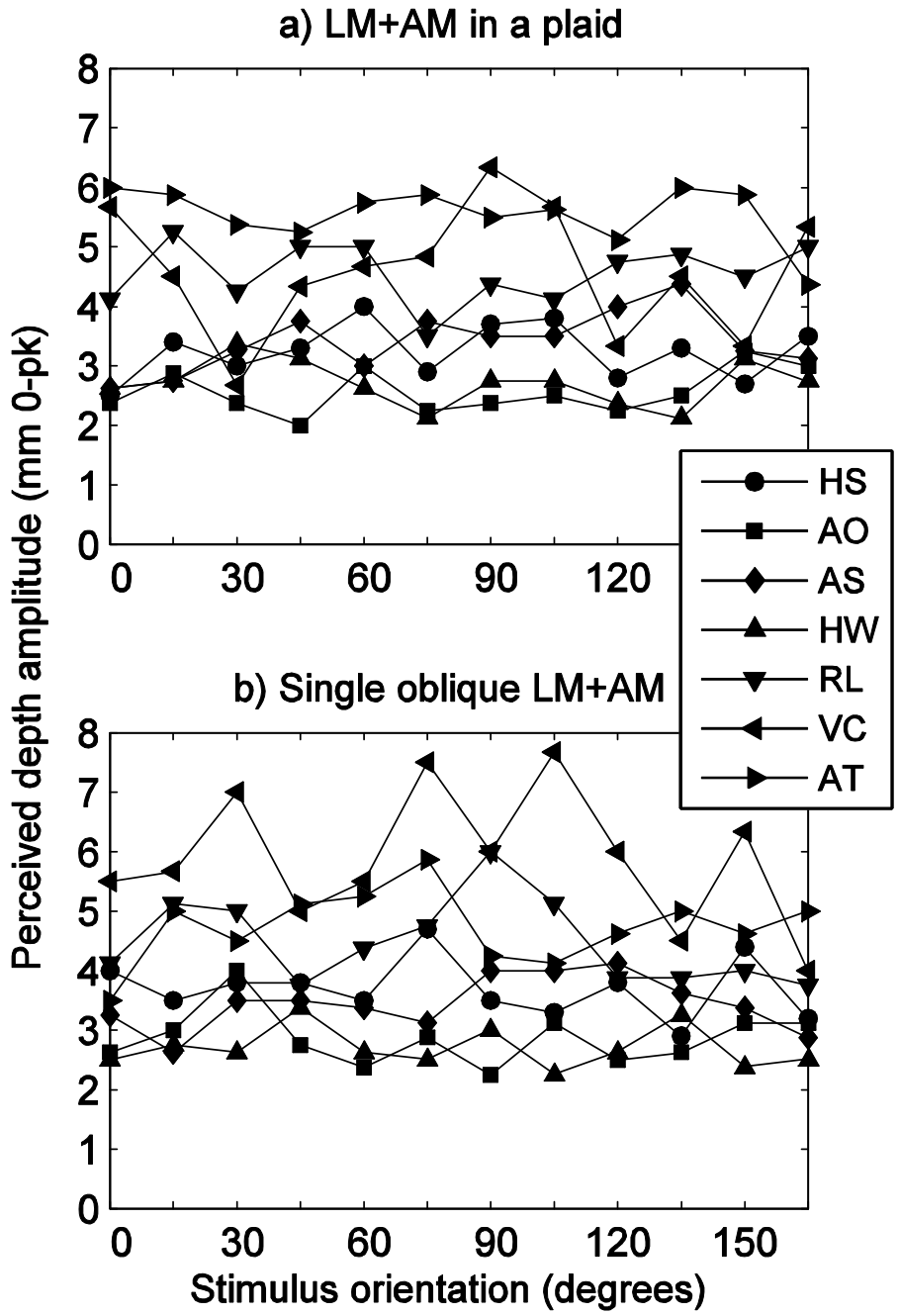

1156
1157 Fig 3



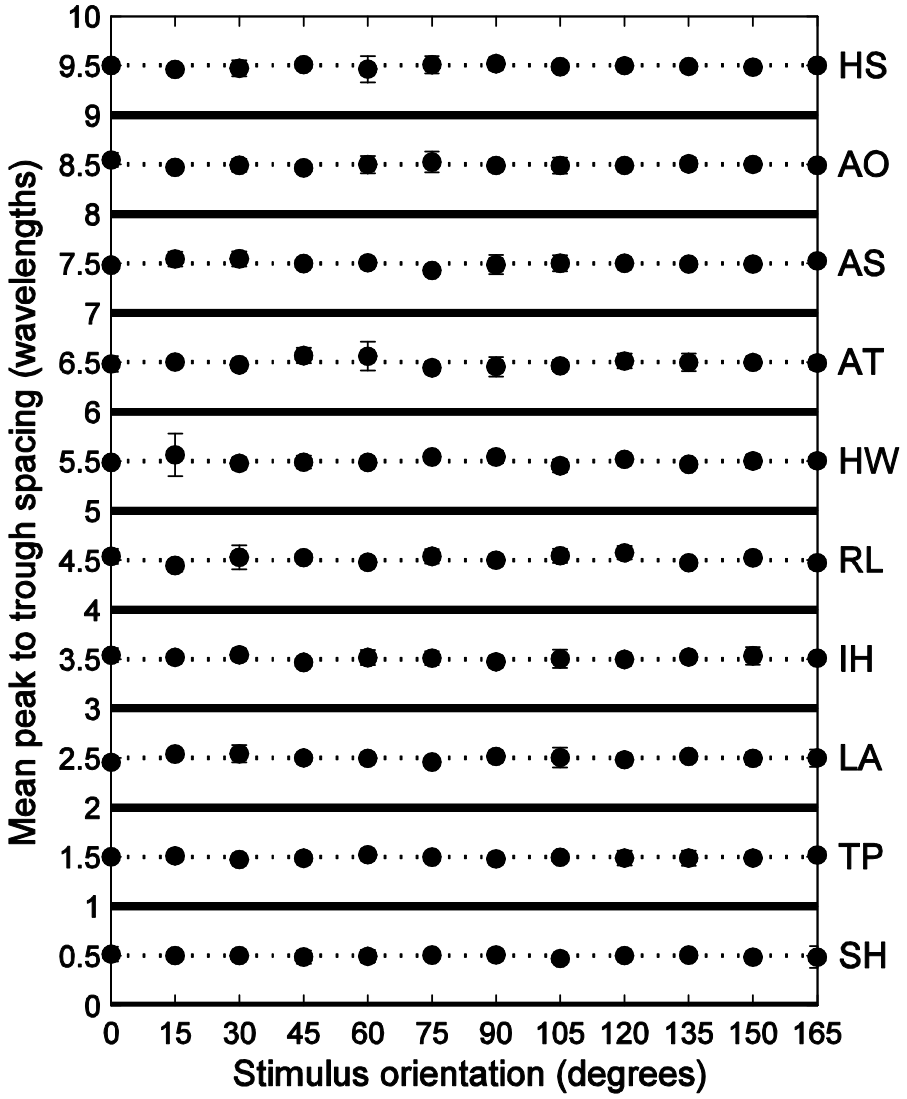

1158
1159 Fig 4



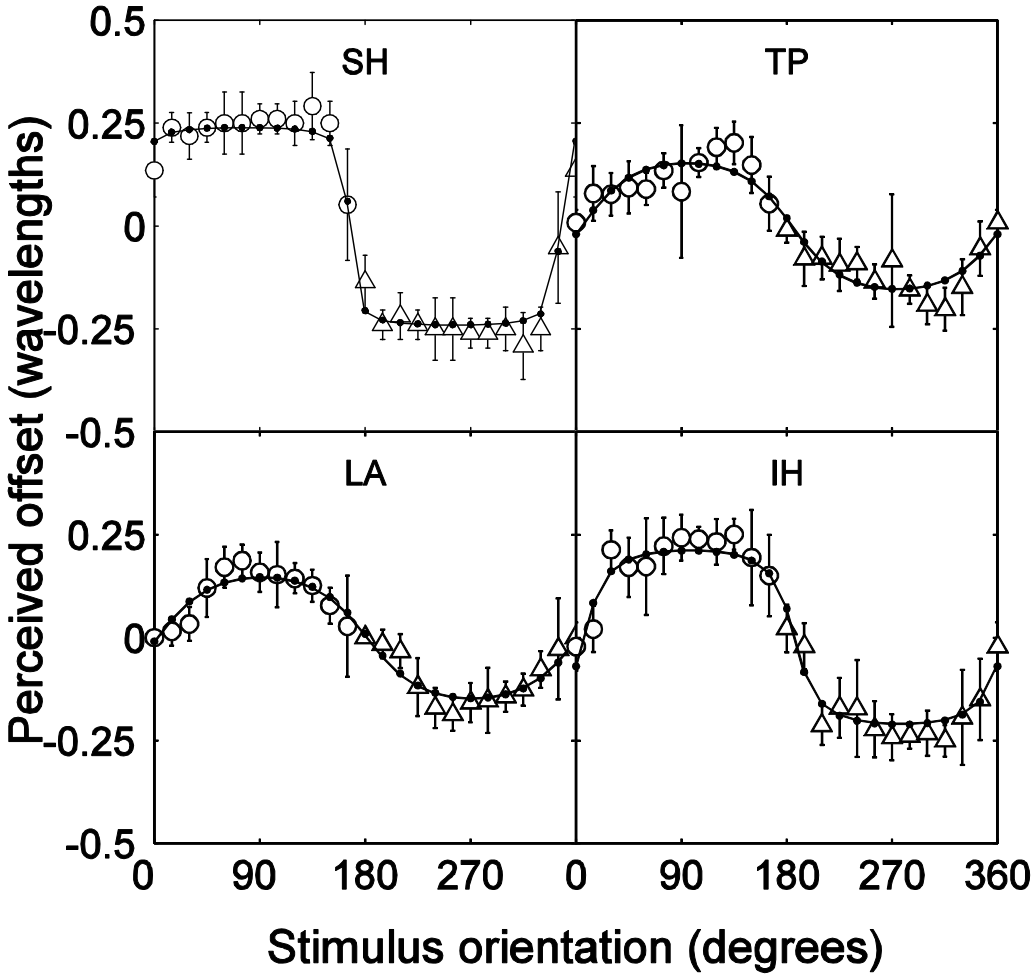

1160
1161 Fig 5



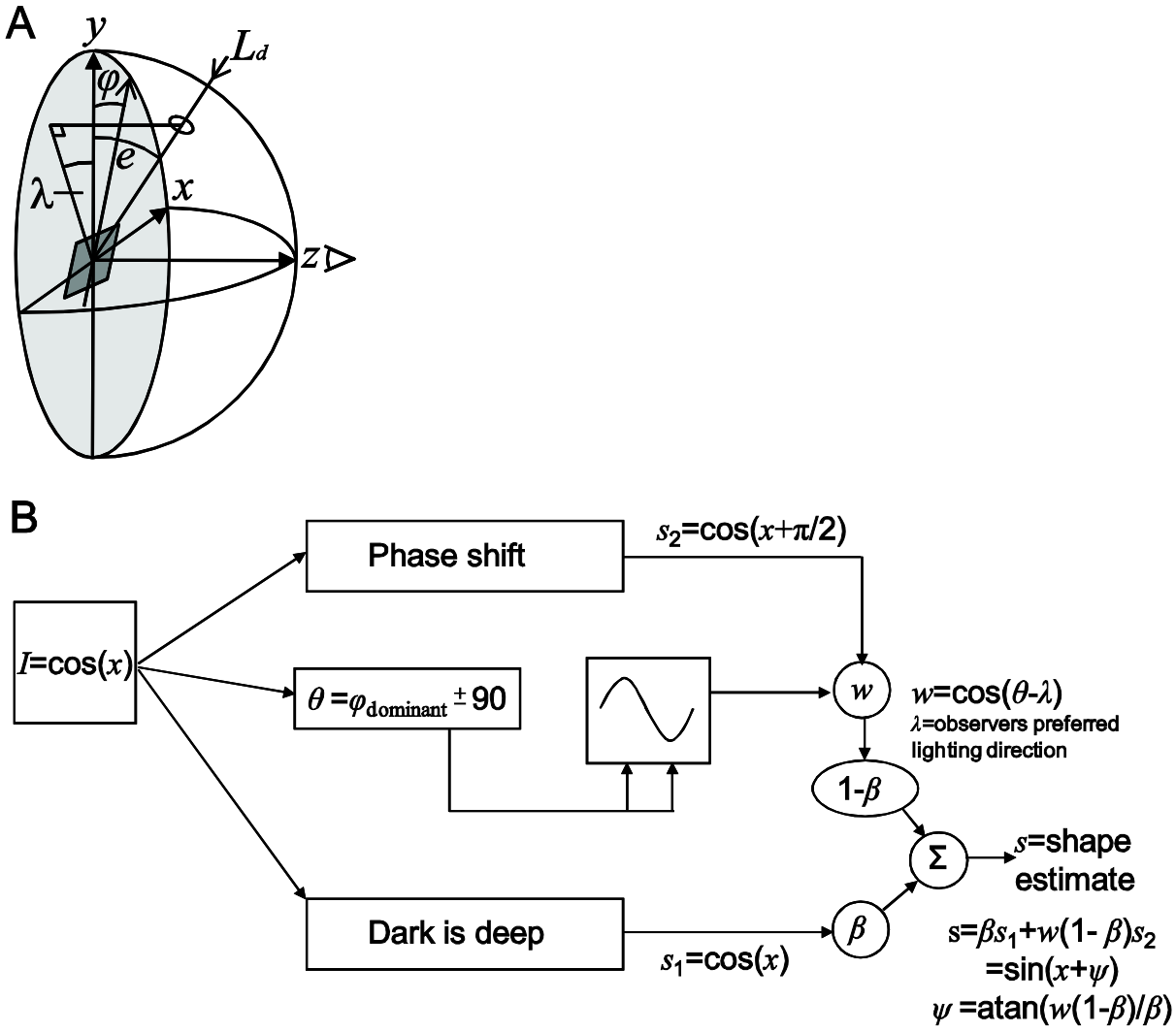

1162
1163 Fig 6




1164
1165 Fig 7




1166
1167 Fig 8




1168
1169 Fig9

Sun and sky: Does human vision assume a mixture of point and diffuse illumination when interpreting shape-from-shading?

Andrew J Schofield, Paul B Rock, & Mark A Georgeson.

Supplementary data.

S1. The relationship between illumination direction and shading profile for sinusoidal surfaces.

One reason for using sinusoidal shading patterns (gratings) is that we know that humans interpret these as sinusoidal surfaces. We have previously shown, in line with Pentland (1988), that sinusoidal luminance signals (single gratings or plaids) produce a convincing percept of a sinusoidally undulating surface, of the same spatial frequency as the luminance profile (Schofield, Hesse, Rock, & Georgeson, 2006). Employing a generative model we now illustrate the point source lighting conditions under which sinusoidal undulations give rise to approximately sinusoidal luminance profiles. We presume that humans adopt one of these lighting interpretations when viewing sinusoidal gratings; if they adopt a point source interpretation at all.

The relationship between sinusoidal surfaces and their luminance profiles under point illumination is illustrated in Figure S1. This figure shows surface profiles (thin red lines) for a vertically oriented sinusoidal surface based on the fronto-parallel plane and undulating in the direction of the line of sight. This surface was rendered under a range of point light sources with the resulting luminance profiles shown by the thick black lines. The polar position of each trace represents the direction and elevation of the light source relative to the surface plane (see legend). The luminance profiles shown are seldom sinusoidal, but when the elevation of the light source is low and its direction oblique to that of the undulations luminance is dominated by a component at the frequency of the undulations (i.e. linear shading; Pentland, 1988). Those parts of the surface most oriented towards the light source have the highest luminance, but do not, in general, correspond to surface peaks (that is, the points closest to the observer for fronto-parallel presentation or more generally points of maximum – convex – surface curvature). In such cases, luminance peaks are offset from surface peaks by ¼ wavelength. When the elevation of the light source is increased or its direction relative to the undulations is more acute, contrast is reduced and a component at twice the undulation frequency is introduced into the luminance profile (i.e. quadratic shading, Pentand 1988)[1]. In such cases the average position of the two luminance peaks is offset by ¼ wavelength from the surface peaks. When the light source is positioned on the line of sight (90° elevation) or directed along the surface ridges (directions 0 and 180°) the frequency-doubled component dominates. Here surface and luminance peaks align but there are additional luminance peaks at the surface minima.

When the elevation of the light source is very low the surface partially occludes itself. Occlusions are worst for obliquely oriented light sources. In many cases occlusion does not shift the position of the luminance peak relative to the surface peak but shifts will occur in extreme cases.

[1] We note that sinusoidal undulations are actually degenerate with respect to the direction of the illumination when expressed in a Cartesian framework. The shape of the luminance profile depends only on the angle between the surface normal and the vector joining the light source to the surface. In our polar representation this single angle is determined by two components that we call elevation and direction. However this distinction is unimportant for two reasons. First, we make no strong claims about the direction of the preferred illuminant in this paper. Second, we did not use the rendered stimuli in this demonstration in our experiments. The purpose of this demonstration is merely to show that sinusoidal surfaces can give rise to approximately sinusoidal luminance profiles over a range of point source lighting conditions and that when this is the case the luminance peaks and physical peaks are offset by ¼ wavelength.

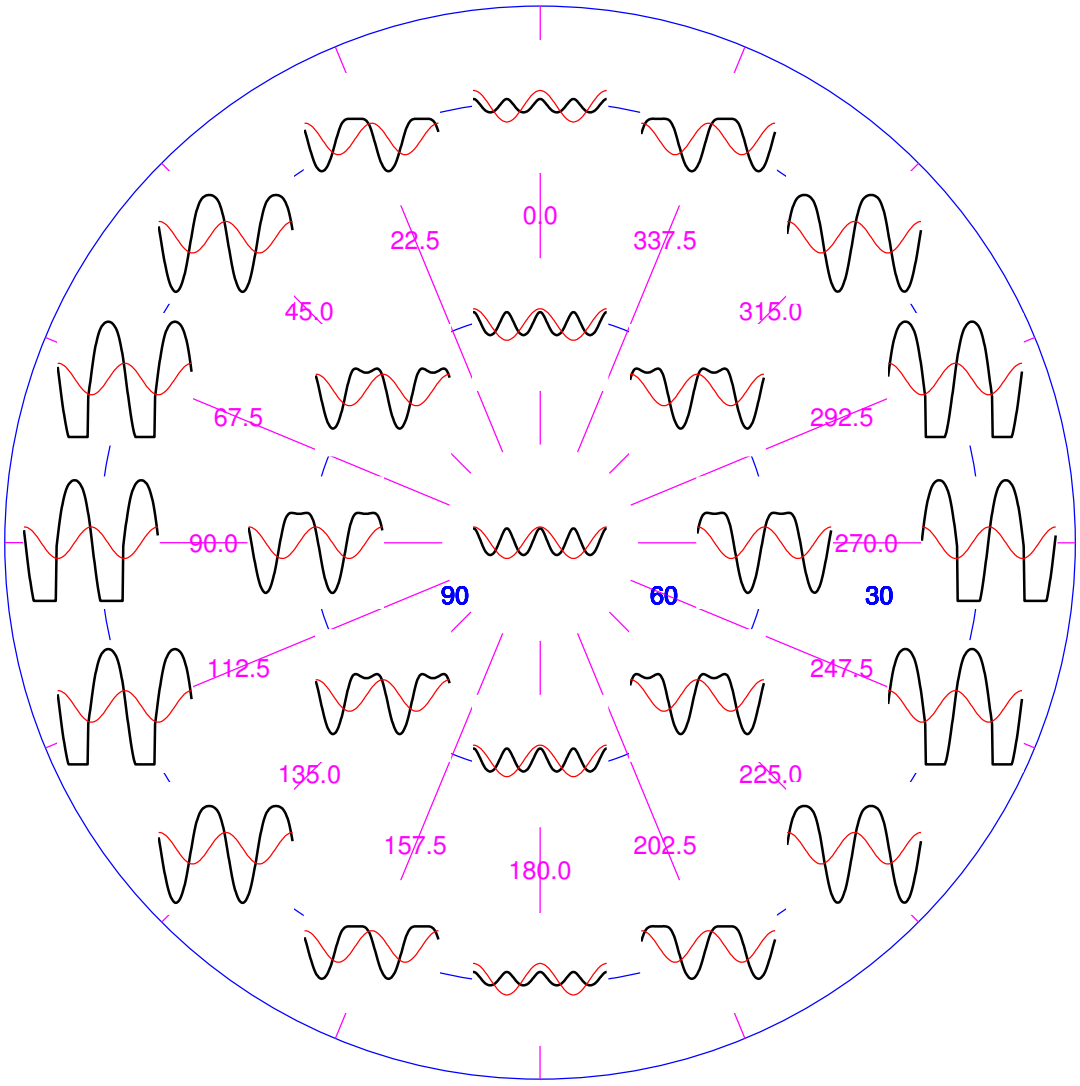



Figure 1. Luminance profiles for a sinusoidal surface under point-source illuminants. Thin red traces show the surface profile; Thick black traces show luminance after rendering the surface under point source illuminants from a range of directions. The surface was sinusoidally corrugated with ridges running top to bottom. Its depth amplitude (mean-to-peak) was 0.12 wavelengths (axes not to scale). The polar location of each pair of traces represents the direction and elevation of the light source relative to the centre of the surface. Polar orientation (magenta radial lines and labels) represents the direction of the light projected into the surface plane while the distance from the centre (blue circles and labels) represents the elevation of the light out of that plane. Note in main Fig 1 radial lines represent surface orientation labelled clockwise; here lighting direction is labelled working anti-clockwise such that relative direction is consistent across the figures. For surfaces based on a horizontal plane direction and elevation should be interpreted as the azimuth and elevation of the 'sun' respectively. However, if the surface is interpreted as based on a vertical plane direction and elevation should be interpreted as follows. When elevation = 90° the light is directly in front of the surface regardless of its direction (central traces) lights with elevations 60° and 30° are progressively closer to the surface plane; direction 0 is above the midpoint of the surface, 90 to its left, 180 below and 270 right. For each sub-plot the x-axis represents position along the surface while the y-axis represents luminance (thick lines) or height (thin lines). Traces were produced in *MatLab* (The Mathworks, Inc) using a Lambertian reflectance model (luminance is proportional to the cosine of the angle between the surface normal and the vector to the light source) with occlusion (luminance is zero when the surface occludes itself from the light source).

The renderings of Figure S1 and main Figure 1 show shading profiles for surfaces with relatively low relief (0.12 of a wavelength, mean to peak depth) illuminated by point light sources with varying direction and elevation. The plots suggest that for such surfaces occlusions are relatively rare for elevations greater than 30° and that while double-crested peaks occur there is a sizable range of elevations and directions for which the quadratic component is relatively small. Thus for low relief surfaces there is an elongated annulus in illumination space for which shading profiles are approximately sinusoidal. In this band – represented best in main Figure 1 – luminance peaks are offset from surface peaks by ¼ wavelength. Figure S1 also suggests that this offset persists even when some occlusions occur and that double-crested peaks are centred on this offset. This analysis suggests that ¼ wavelength offsets are common for such stimuli. We investigated this notion further by rendering surfaces under every possible illuminant in the space depicted in Figure S1 (sampled on a 181x181 position grid). We then measured a number of properties of the resulting shading images. We estimated the location of the dominant peak in the stimuli by measuring the phase of the frequency component with the same frequency as the surface (the fundamental frequency). This is represented by hue in Figure S2a with purple and green equal to ¼ wavelength offsets in the positive and negative directions respectively. We also measured the degree to which the shading profiles were sinusoidal by taking the ratio of the amplitude of the fundamental to all of the higher order components combined. This is represented by saturation in Fig S2a. Finally, we measured the degree of occlusion by simply measuring the proportion of 'pixels' on the surface that can 'see' the illuminant, 1= no occlusions, 0=surface totally occluded. This is represented by intensity in Fig S2a.

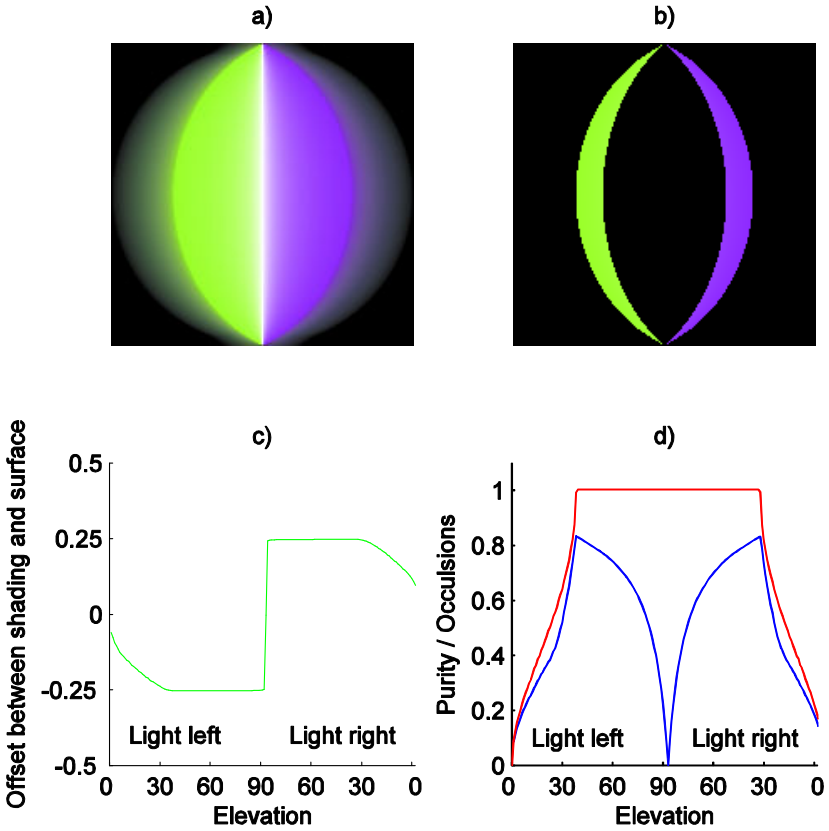



Figure S2. a) Representation of the illumination space. Hue corresponds to the offset between surface peaks and the fundamental component of the luminance waveform – normally the dominant peak. Saturation represents the 'purity' of the shading waveform given by the amplitude of the fundamental divided by the total amplitude of all other components. Intensity represents the proportion of surface points that can see the light source. b) as (a) but only those light locations that produce neither occlusion nor double peaks are coloured. c) cross section through (a) showing offsets. d) cross sections through (a) showing purity (blue) and 'occlusions' (red) where 'occlusions' records the proportion of surface positions that can *see* the light source.

Fig S2b shows offsets (hue), and purity (saturation), for only those lighting directions that produce no occlusions at all and which contain no evidence of a double peak.

Fig S2c shows offsets for a cross section through the midline of Fig 1a.
Fig S2d shows purity (blue) and occlusions (red) for the same cross section as Fig S2c.

It can be seen that the peak of the fundamental component of the shading profile is almost always ¼ wavelength offset from the surface peaks. This breaks down only when either the quadratic component in the shading profile begins to dominate or when the degree of occlusion is large. We conclude that for shading stimuli that are dominated by linear component shading - surface offsets will generally be ¼ wavelength. Importantly our sinusoidal stimuli can provide a reasonably good match to the shading profiles produced from surfaces whose half-height amplitude is 0.12 of their wavelength – that is, the haptic surfaces of Experiment 1.

The depth of the undulations represented in Figures S1 and S2 and main Figure 1 is not accidental but (as a proportion of wavelength) is the same that used for the haptic stimuli of Experiment 1. These figures clearly show that sinusoidal luminance profiles are a physically plausible representation of sinusoidal surfaces that are configured as in Experiment 1 and rendered under point light sources. Further, this illustration strongly suggests that in perceiving sinusoidal gratings as sinusoidally undulating surfaces humans should, if they adopt a point source illumination model, also perceive the surface peaks to be offset by ¼ wavelength from the luminance peaks. We assert this because, in the generative case, such offsets are ubiquitous whenever sinusoidal surfaces produce even approximately sinusoidal luminance profiles. We discuss models for how people might perceive luminance peaks to be offset from surface peaks by ¼ wavelength in Section 1.2 of the main paper.

We conclude that point source illumination of a low-relief sinusoidal surface produces approximately sinusoidal luminance profiles over a range lighting directions. Except for frontal lighting, lighting directed in the direction of surface ridges, and very oblique lighting, luminance profiles have the same fundamental frequency as the surface itself but their peaks are offset from the surface peaks by ¼ wavelength.

## S2. Control Experiment 1: Do observers see sinusoidal luminance patterns as sinusoidal surfaces?

### S2.1 Rationale

Our experiments rely on the observation that people perceive sinusoidal luminance gratings as sinusoidal surfaces. Our previous results (Schofield, et al., 2006) suggest that this is indeed the case, as do the results presented by Pentland (1988). However, neither of these studies used the same stimulus presentation as we do here. In particular they used vertically presented stimuli rather than stimuli inclined at 45°. Further, neither study formally tested the hypothesis that the shape of the perceived surface is sinusoidal versus some other profile. This control experiment addresses these issues using a gauge figure task to assess perceived gradient at points along each sinusoidal modulation and thence to estimate perceived surface shape.

### S2.2 Method

Stimuli (see Fig S3) were similar to those of experiments 1 and 3 and consisted of either LM+AM / LM-AM plaids or single sinusoidal LM+AM gratings applied to samples of isotropic binary noise. We tested six different orientations (0, 30, 60, 90, 120, 150°) at eight different phases (0, 45, 90, 135, 180, 225, 270 and 315 degrees of phase). Each orientation / phase combination was tested 10 times. Stimuli were presented in the ReachIn haptic workstation described in Experiment 1 and thus on a plane tilted backwards at 45. However, observers looked down at the stimuli at a similar angle so the stimulus plane was fronto-parallel. A gauge figure, comprising a 2D rendition of an unfilled ellipse with an orthogonal stick (see Fig S2) was placed at the centre of each image and observers were asked to adjust the apparent orientation of this figure until it appeared to lie on the surface. By changing the phase of the sinusoid we effectively moved it under the gauge figure allowing us to measure gradients at different positions on the surface. The gauge figure could only be adjusted in the direction orthogonal to the sinusoidal modulation under test. That is, for a vertical sinusoid the gauge figure could only be made to point to the left or right. The initial orientation of the gauge figure was randomised within the range ±85° degrees at the start of each trial. The phase of the untested LM-AM component of the plaid stimuli could be either 90 or 270 degrees such that the gauge figure was always at a zero crossing for this cue.

Six naive observers took part in the study, three each completing the plaid and single oblique conditions. None had participated in our previous experiments. All other experimental details were as described in Experiment 1 of the main study.

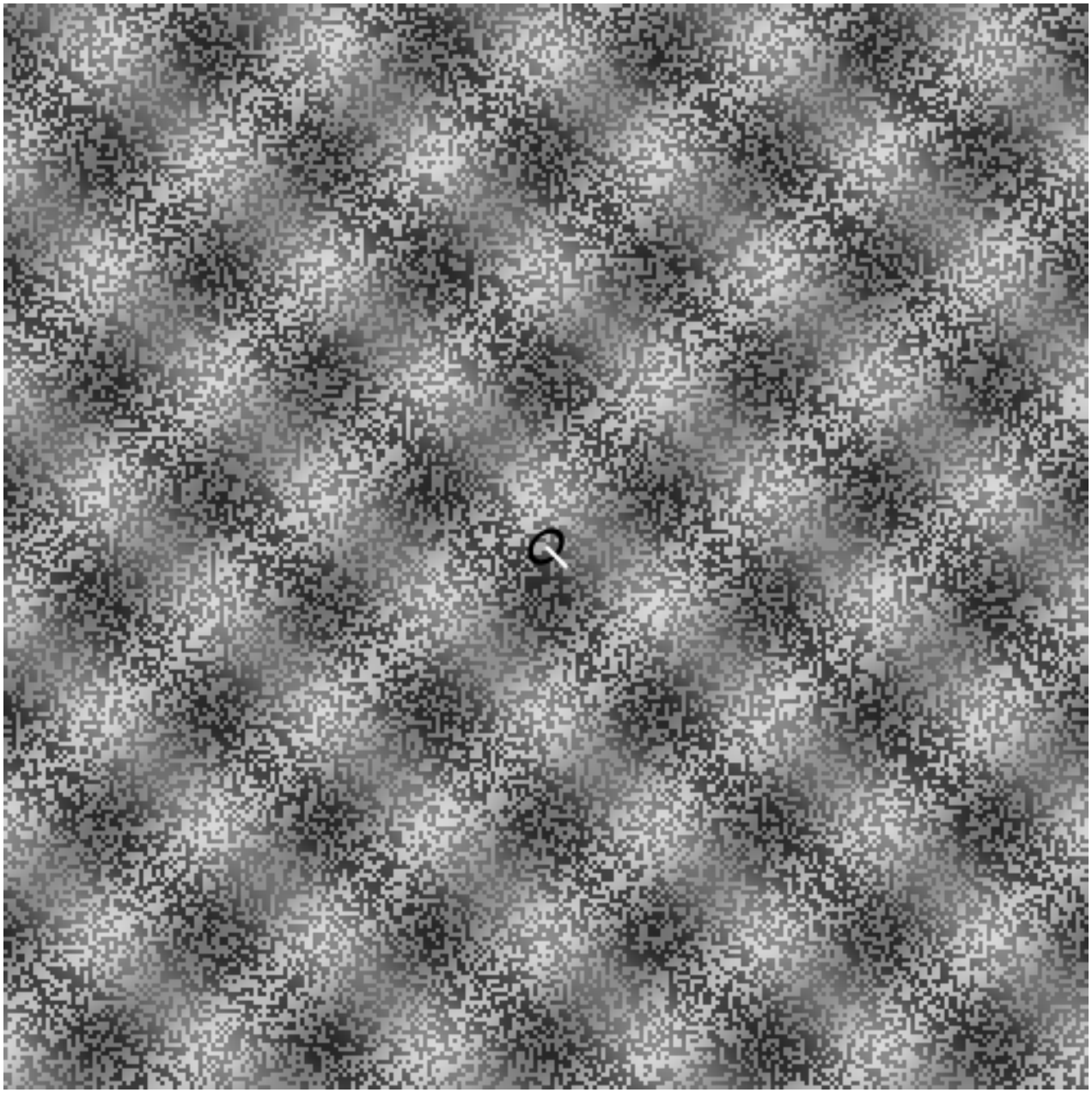


Figure S3a. Example plaid stimulus from control experiment 1 showing gauge figure. Noise contrast has been exaggerated to aid visualisation.

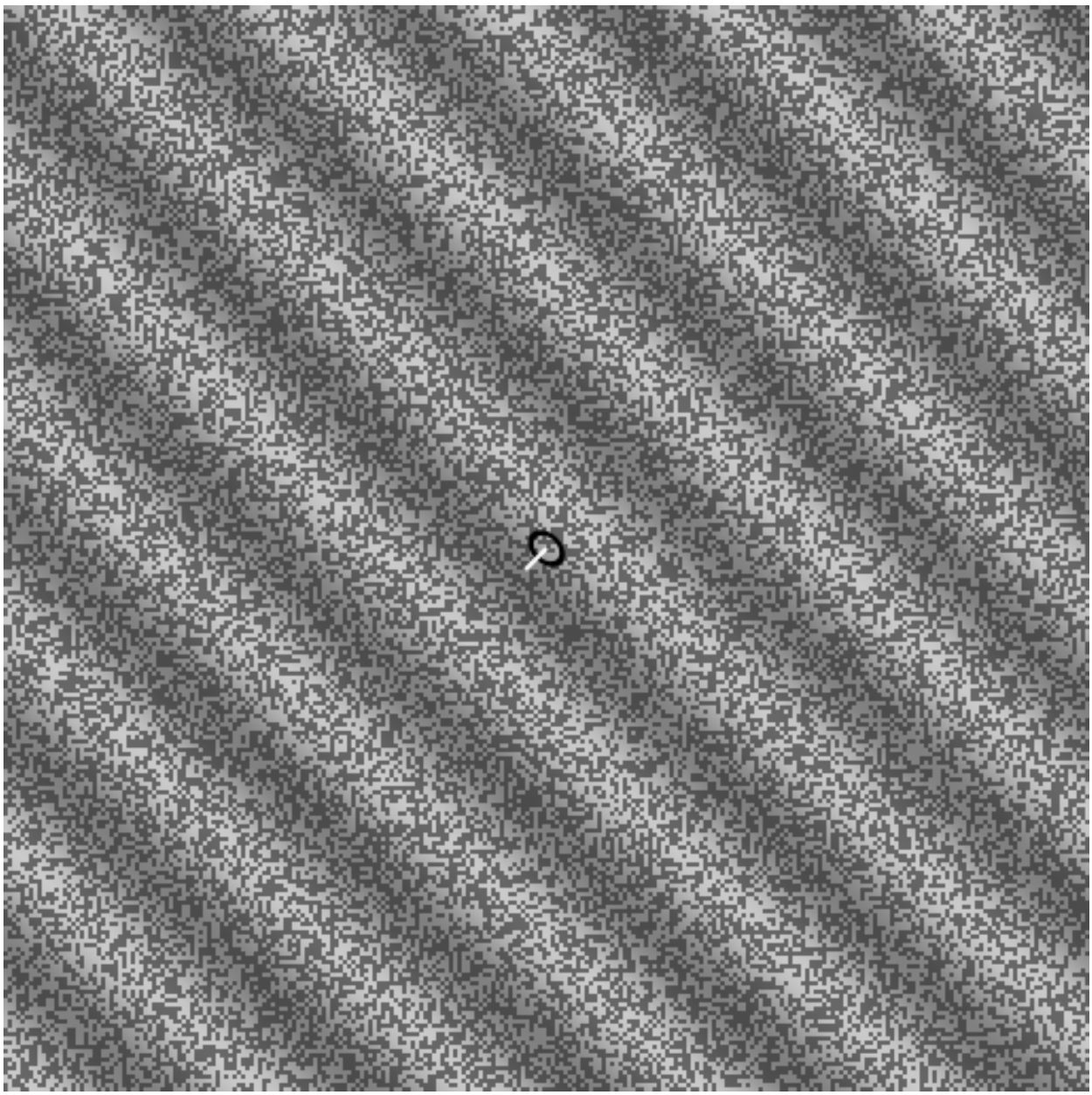


Figure S3b. Example single oblique stimulus from the control experiment showing gauge figure. Noise contrast has been exaggerated to aid visualisation.

S2.3 Analysis

We took the mean of the ten gradient estimates for each orientation / phase combination. We then collated the data for all phases at a given orientation and fit these data with three plausible gradient profiles representing 4 possible surface interpretations. The functions used were: sinusoid, square-wave, and saw-tooth. Once integrated to form perceived surfaces these gradient functions correspond to the following surface profiles: sinusoidal gradient profiles integrate to produce sinusoidal surface profiles with a phase shift of 90 degrees (1/4 wavelength), square-wave gradient profiles integrate to form triangle wave surfaces with a surface peak located at each falling edge in the gradient profile, negative going saw-tooth ramps integrate to produce a series of humps with sharp valleys (humps peak as the saw-tooth ramp crosses y=zero), positive saw-tooth ramps produce sharp ridges and shallow valleys (ridge peaks occur at the abrupt fall in the saw-tooth). We varied the phase (position), amplitude and dc level of each gradient profile in order to obtain the best (least-mean-square) fits. Fits were obtained using MatLab's *fminsearch* function and were compared by calculating the correlation between the predicted and actual gradients for each fit; $R^2$ values are reported. We were thus able to compare 4 candidate surface profiles: sinusoidal, triangular, smooth-humps and sharp-ridges. Although by no means exhaustive, we feel that these four profiles represent good candidates for the perceived surface in response to our sinusoidal shading waveforms given that they are universally, but informally, described as 'undulating corrugations' by our participants.

S2.4 Results

Figure S4 shows a typical dataset and fitted curves for one participant at one orientation. Table S1 shows the $R^2$ values for all six observers, six orientations and three fitted gradient profiles. $R^2$ values for the sine wave profiles are generally quite high (typically > 0.8). The sinusoidal model provided the best fit in 34 of 36 cases: at all orientations for four observers and at all but one orientation for the remaining two observers. We conclude that the sinusoidal gradient profile provides a better fit to the data than either the square-wave or saw-tooth profiles. We also conclude that observers see sinusoidal shading patterns as sinusoidal surfaces most of the time. We note that there was very little variation in the amplitude of the best fit gradient profiles with orientation. This result suggests that the surfaces are seen as having about equal depth amplitude at all orientations: confirming the results of Experiment 2.

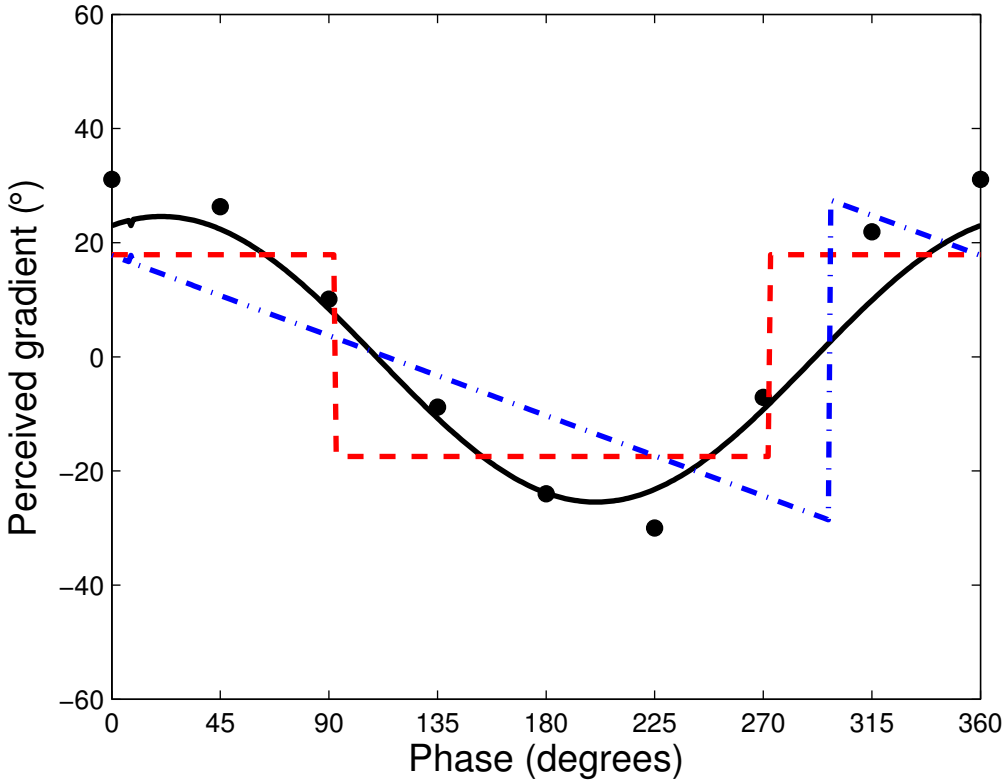



Figure S4. Gradient profile fits for an example dataset from one observer. Symbols, raw data; black solid line sinusoidal gradient profile; red dashed line, square-wave gradient profile; blue dot-dash line, saw-tooth gradient profile.

| | CC | | | DD | | | HAW | | |
|---|---|---|---|---|---|---|---|---|---|
| Orientation | Sine | Square | Saw tooth | Sine | Square | Saw tooth | Sine | Square | Saw tooth |
| 0 | .97 | .76 | .63 | .99 | .79 | .66 | .98 | .72 | .63 |
| 30 | .7 | .77 | .76 | .85 | .69 | .56 | .82 | .78 | .73 |
| 60 | .78 | .63 | .41 | .93 | .76 | .56 | .91 | .75 | .71 |
| 90 | .81 | .67 | .6 | .9 | .69 | .52 | .95 | .8 | .74 |
| 120 | .84 | .65 | .58 | .91 | .67 | .5 | .89 | .84 | .81 |
| 150 | .82 | .69 | .46 | .8 | .67 | .64 | .86 | .83 | .68 |

Table S1a. $R^2$ values for the three gradient profiles for each observer at each orientation in the plaid condition.

| | GM | | | KC | | | XJ | | |
|---|---|---|---|---|---|---|---|---|---|
| Orientation | Sine | Square | Saw tooth | Sine | Square | Saw tooth | Sine | Square | Saw tooth |
| 0 | .87 | .74 | .83 | .89 | .14 | .46 | .84 | .65 | .62 |
| 30 | .79 | .73 | .62 | .66 | .17 | .51 | .78 | .7 | .65 |
| 60 | .88 | .83 | .67 | .76 | .75 | .51 | .81 | .59 | .44 |
| 90 | .95 | .75 | .61 | .58 | .58 | .27 | .89 | .69 | .47 |
| 120 | .95 | .8 | .63 | .7 | .71 | .41 | .88 | .72 | .56 |
| 150 | .85 | .77 | .21 | .71 | .62 | .35 | .78 | .64 | .6 |

Table S1b. $R^2$ values for the three gradient profiles for each observer at each orientation in the single oblique condition.

As noted in Section S3.3 it is possible to estimate the position of the perceived surface peak relative to the luminance peak from the model fit parameters. We computed such estimates based on whichever of the three gradient profiles produced the better fit in each condition. The resulting inter-peak offsets are shown in Fig S5 which has the same format as Fig 3 in the main paper. Inter-peak offsets were fit using the mixed illumination model described in the main paper and the resulting fit parameters are shown in Table S2. These are consistent with those shown in the main paper.

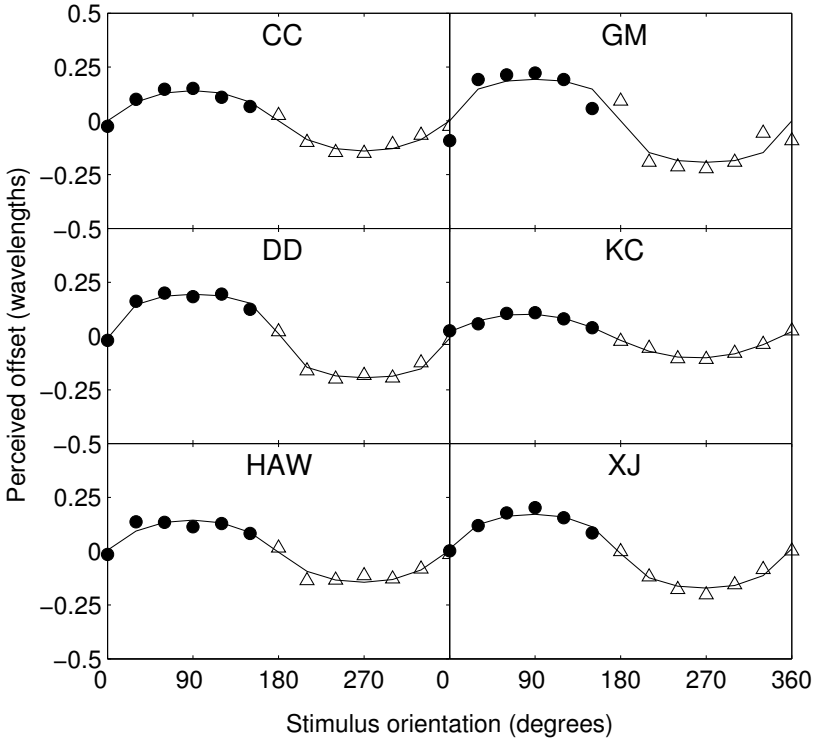



Figure S5. Perceived offsets as a function of stimulus orientation for the three participants in control experiment 1. Circles represent recorded data, triangles interpolated data. Lines represent model fits. For details of the data interpolation and modelling processes, see the main paper.

| Stimulus | Person | SSE | β | Preferred light source (λ) |
|---|---|---|---|---|
| Plaid | CC | .002 | 0.45 | 0.7 |
| Plaid | DD | .002 | 0.27 | -1 |
| Plaid | HAW | .003 | 0.44 | 2.8 |
| Single | GM | .018 | 0.247 | -5.8 |
| Single | KC | .000 | 0.57 | 11.7 |
| Single | XJ | .002 | 0.35 | 2.9 |

Table S2. Model fit parameters. Columns show the sum of squared errors between the modelled and observed inter-peak offsets, the weight (β) applied to the diffuse interpretation, and the observers' preferred lighting direction (the one giving the most positive weight for the point source interpretation). Lighting direction is given in degrees; positive = anti-clockwise shift from vertical.

S2.5 Discussion

The results of this control experiment confirm that observers tend to see sinusoidal shading patterns as sinusoidal surfaces. This finding adds further validation to our approach of asking observers to match haptically presented sine waves to our visual stimuli. The haptic method has the advantage that inter-peak offsets can be assessed more quickly and more directly than in the case with the gauge figure method. Nonetheless the derived inter-peak offsets and resulting model fits for this control are consistent with those obtained in the main study.

S.3 Model equivalence.

The similarity of the two model fits described in the main paper suggest that they two models are mathematically equivalent at least for the case of sinusoidal stimuli. We now demonstrate that this is indeed the case. Basing Model A offsets profiles only on the Fourier component at the frequency of the physical surface is equivlant to simplifying Eqn 5 keeping only those terms at that frequency:

$$L \propto (1-\gamma)\left(-0.12\sin(x)\cos\left(\frac{\pi}{2}+\varphi+\lambda\right)\sin\left(\frac{\pi}{2}-e\right)\right)+\gamma\left(0.065(\cos(x))\right) \qquad \text{Eqn 6}$$

Recalling that e is constant, λ constant for a given fit and φ the direction of the surface corrugations the lumaince profile for a given φ becomes the sum of two orthogonal sine waves with the same frequency and different amplitudes: sin(x) with amplitude $-0.12(1-\gamma)\cos(\pi/2-\varphi+\lambda)\sin(\pi/2-e)$ and cos(x) with amplitude $0.065\gamma$. Thus the phase of Eqn 6 is given by:

$$\arctan\left(-\frac{0.12(1-\gamma)\cos(\pi/2+\varphi+\lambda)\sin(\pi/2-e)}{0.065\gamma}\right) \qquad \text{Eqn 7}$$

combining all the constants and given that θ in Model B is equal to φ-π/2 and that cos is symmetric about zero we can rewrite eqn 7 this as:

$$\arctan\left(\frac{0.12\sin(\pi/2-e)(1-\gamma)\cos(\theta-\lambda)}{0.065\gamma}\right)$$



Citation
Sun P., & **Schofield** A.J. (2012) Two operational modes in the perception of shape-from-shading revealed by the effects of edge information in slant settings. Journal of Vision, 12 (1): 12. doi:10.1167/12.1.12