

*promoting access to White Rose research papers*



**Universities of Leeds, Sheffield and York**  
**<http://eprints.whiterose.ac.uk/>**

---

This is an author produced version of a paper published in **Perception**.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/42967/>

---

**Published paper**

Stone, JV (2011) *Footprints Sticking Out of the Sand (Part II): Children's Bayesian Priors For Shape and Lighting Direction*. *Perception*, 40 (2). 175 -190.  
ISSN 0301-0066

<http://dx.doi.org/10.1068/p6776>

---

Footprints Sticking Out of the Sand (Part II):  
Children's Bayesian Priors For Shape and Lighting Direction

Target journal: Perception

File: bayes\_chilld\_Jan2011\_v23\_sent.tex

JV Stone

Department of Psychology, University of Sheffield, England, S10 2TP.

Corresponding author: j.v.stone@shef.ac.uk

Key words: Perception, Bayes, lighting direction,  
developmental priors, shape convexity, shape from shading.

Short title: Children's Bayesian Priors

February 4, 2011

## Abstract

The shading information in images that depict surfaces of 3D objects cannot be perceived correctly unless the direction of the illuminating light source is known, and, in the absence of this knowledge, perception in adults is consistent with a light-from-above Bayesian prior assumption. In order to investigate if children make use of a similar assumption, 171 children between the ages of 4.6 and 10.8 years were tested using 20 images containing shading information, where the shape depicted in each image can be perceived as either convex or concave. Each child's Bayesian prior probability that light comes from above was estimated, and (assuming that decision-noise is approximately the same in all children) regression analyses revealed a significant increase in this prior probability of 0.034-0.035 per year, and predict a neutral prior (ie 0.5) at 1.6 years for naturalistic picture stimuli and 3.6 years for abstract symbol stimuli. Additionally, each child's prior probability for perceiving shapes as being convex/concave was estimated, and was found to be close to a neutral value of 0.5 for all ages. Together, these results indicate that children have a neutral prior for shape convexity, and that their prior probability for lighting direction gradually shifts towards an adult-like prior value as they grow older. Finally, the status of these one-dimensional priors is discussed in relation to marginal distributions of high-dimensional priors implicit in the statistical structure of the physical world.

## 1 Introduction

Perception involves recovering the three-dimensional structure of a scene from a two-dimensional retinal image. Unfortunately, the process of projecting a scene onto the retina discards information about the three-dimensional structure of that scene. This makes it impossible, in principle, to recover the scene structure from a retinal image alone, making perception a classic example of an ill-posed problem (Poggio et al. (1985)). It is therefore necessary for the visual system to rely on extra information, in the form of constraints, assumptions, or Bayesian priors.

One assumption which adults adopt is that light comes from above (Rittenhouse (1786); Brewster (1826); Metzger (2009); Von Fieandt (1949); Berbaum et al. (1983, 1984); Sun and Perona (1998); Mamasian and Landy (2001); Stone et al. (2009)). For example, the images in Figure 1 can be interpreted either as convex or as concave, depending on the direction from which light is assumed to originate. However, the finding that perception of similarly ambiguous shaded figures can be modified by experience in adults

(Adams et al. (2004)), suggests that adults may have learned to interpret such figures during development. In contrast, chickens may be immune to the effects of experience (Hershberger (1970)), although earlier experiments provide some indication that exposure to lighting-from-below from birth to 7 weeks causes chickens to interpret stimuli as if light comes from below (see Hess (1950); Woodworth and Schlosberg (1954), p463).

Brewster (1826) noted that children do not always perceive ambiguous shaded figures as adults do. More recently, it has been reported that 3-8 year olds have an increasing tendency to interpret a single dome-like shaded stimulus as if light comes from above (Yonas et al. (1979)),<sup>1</sup> and that this applies to ‘polo mint’ stimuli in 4-12 year olds (Thomas et al. (2010)), and also to abstract symbols and naturalistic stimuli in 4-11 year old children (Stone and Pascalis (2010)). There is evidence that 7 month old infants (but not 5 month olds) perceive a picture of a dome-like shaded stimulus as if light comes from above (Granrud et al. (1985)). These results are consistent with those recently reported in (Tsuruhara et al. (2009)); using stimuli in which depth was implied by one of two different cues (shading or texture), it was found that 6-7 month old infants showed transfer-across-depth-cues, while 4-5 month-old infants did not. Both of these studies suggest the existence of a light-from-above prior by the age of 7 months.

Here, the data presented in (Stone and Pascalis (2010)) are used to investigate if the developmental changes reported above are contingent on underlying changes in a Bayesian prior for lighting direction and shape convexity.

**Place Figure 1 about here.**

## **2 The Light-From-Above Prior and the Convexity Prior**

The light-from-above (LFA) prior can be considered as the propensity of a child to assume that light comes from above *in the absence of a stimulus*. As the stimuli used here have light coming only from above or below, the analysis of each child’s data was performed using a model which included only these two lighting directions. The result of assuming each stimulus has only one of two possible lighting directions is that the ‘all around the clock’ prior distribution collapses to a prior ‘distribution’ model with only two

---

<sup>1</sup>However, it is not known if infants were repeatedly rewarded with praise for reaching toward the same apparently convex shape (which was presented 16 times), which accounts for the results if infants simply learned to reach for the shape that was lighter at the top.

lighting directions (above/below). This means that the propensity of each observer to assume light comes from either above or below can now be defined in terms of a single number  $\rho$  between zero and one, which is the relative probability that an observer assumes light comes from above. For example, an observer with a prior of  $\rho = 1.0$  always assumes light comes from above, whereas  $\rho = 0.5$  implies an equal propensity for assuming light comes from above and below.

Similarly, the convexity prior can be considered as the propensity of a child to assume that objects are convex in the absence of a stimulus. As the stimuli used here can be perceived only as convex or concave, this effectively limits the number of shapes under consideration to two. This means that the propensity of each observer to assume convexity or concavity can be defined in terms of a single number  $\gamma$  between zero and one, which is the relative probability that an observer assumes convexity.

### 3 Results

#### Place Figure 2 about here.

The stimuli were 5 naturalistic pictures and 5 geometric embossed symbols, with 2 examples shown in Figure 1. Each stimulus was presented twice, the right way up, and upside-down, making a total of 20 stimuli. The mean ages of children in each bin were [4.8, 5.5, 6.5, 7.4, 8.4, 9.8, 10.4] years, the corresponding numbers of children that contributed to data in each bin were [16, 32, 39, 29, 21, 12, 22], and the mean probability of a ‘correct’ response<sup>2</sup> within each age group is [0.572, 0.589, 0.686, 0.722, 0.745, 0.783, 0.802]. A full account of the stimuli, methods, participants and initial data analysis are given in (Stone and Pascalis (2010)). In that paper, we reported a significant relationship between age and children’s tendency to interpret shading information as if light comes from above, but no significant change in children’s tendency to perceive stimuli as being convex. Here, we report a significant increase in the value of a light-from-above prior with age, and a non-significant change in the convexity prior with age.

**Changes in the light-from-above prior with age:** The Bayesian light-from-above prior for each child was estimated using a modified form of the method described in Stone et al. (2009), (see Appendix). For symbol stimuli, the mean light-from-above prior is  $\bar{\rho}_{sym} = 0.63$  (sem= 0.013), and for pictures the mean is  $\bar{\rho}_{pic} = 0.70$  (sem= 0.012), both of which are significantly different from a neutral prior of 0.5 ( $p < 0.001$ ).

---

<sup>2</sup>A ‘correct’ response is defined here as a response consistent with overhead light.

A separate regression analysis was also performed on the picture and symbol stimuli, see legend of Figure 2. For both symbol and picture stimuli, the slopes of the fitted regression lines are non-zero ( $p < 0.05$ ), which suggests that the propensity of each child to assume that light comes from above increases throughout childhood. The two fitted regression lines have slopes that are not significantly different ( $p > 0.05$ ), so that the light-from-above prior increases at about the same rate for symbol and picture stimuli.

These regression analyses predict that the lighting direction prior has a value of  $\rho = 0.5$  at 3.63 years for symbol stimuli, and 1.61 years for picture stimuli, which suggests that children make no particular assumption about lighting direction below these ages for these type of stimuli. A paired t-test comparing the 7 means plotted in Figure 2a with those in Figure 2b yielded a significant difference ( $t = 7.18, p < 0.001$ ), so the light-from-above prior for picture stimuli is significantly greater than the light-from-above prior for symbol stimuli.

## Place Figure 3 about here.

**Changes in the convexity prior with age:** The method of analysis also provides an estimate of the convexity prior  $\gamma$ , which underlies the probability of a *convex* response, as shown in Figure 3. The mean convexity prior for symbol stimuli is  $\gamma_{sym} = 0.446$  (sem=0.013), which is significantly smaller than a neutral convexity prior of  $\gamma = 0.5$  ( $p < 0.01$ ). For picture stimuli  $\gamma_{pic} = 0.509$  (sem=0.011), which is not significantly different from  $\gamma = 0.5$ . However, each of these two stimulus-specific means is based on seven age-specific means, some of which are significantly different from each other (as indicated by the error bars in Figure 3), so that collapsing across age to obtain the two stimulus-specific means should be treated with caution.

A separate regression analysis for the changes in the convexity prior with age yielded non-significant fits for both symbol and picture stimuli (see legend of Figure 3), and neither slope is significantly different from zero ( $p > 0.05$ ).

## 4 Discussion

The main findings from this paper and its partner (Part I), are threefold. 1) As children grow older, they have an increasing tendency to interpret a range of ambiguous naturalistic picture and symbolic stimuli as if light comes from above. 2) Estimates of the intrinsic lighting-direction prior suggest that children's assumption regarding lighting direction evolves throughout childhood towards an increasingly adult-like light-from-

above prior. 3) Estimates of the intrinsic shape-convexity prior suggest that children are essentially neutral with respect to shape convexity. The findings (1) and (2) may appear to be very similar, but they are qualitatively different. This is because, in essence, the phenomenological findings of (1) depend upon the (now un-)hidden variable, namely the light-from-above prior of (2).

The results presented here could be interpreted either in terms of learning the statistical structure of the physical world (eg light tends to come from above), or in terms of a pre-programmed developmental trajectory. The experiment presented here does not permit these two alternatives to be discriminated.

**The Convexity Prior:** Results reported here provide evidence for a neutral ( $\rho = 0.5$ ) convexity prior (a result that is in agreement with previous estimates in adults (Stone et al. (2009))), and no indication that this changes consistently with age. However, a smaller reaction time for convex shapes than for concave shapes has been cited as indirect evidence for a non-neutral convexity prior in adults (Kleffner and Ramachandran (1992)), and Liu and Todd (2004) concluded that a strong bias for convexity exists alongside a weaker bias for overhead lighting direction, but these studies did not have access to methods that could measure the priors for convexity or lighting direction. More recently, it was claimed in (Thomas et al. (2010)) that a convexity prior dominates perceptual interpretation of shaded figures in young children, but that a light-from-above prior dominates later and in adulthood. However, it is difficult to interpret overt responses in terms of priors in the absence of a rigorous method for estimating those priors. For example, the raw data for picture stimuli reported in Figure 5a of (Stone and Pascalis (2010)), give the appearance of a non-neutral prior for convexity (and the proportion of convex responses is significantly above chance), but these data have been shown to be consistent with a *neutral* (ie 0.5) convexity prior here.

The relationship between the observer responses and the estimated priors for convexity and lighting-direction is systematic, but non-trivial. Recovering priors from observer's data is analagous to solving the following algebraic problem: find the value of the prior  $x$  given that an observer's response  $y = f(x \times z/w)$ , where neither  $x$  (the prior), nor  $z$  (the likelihood), nor  $w$  (the evidence) are known, and only the general (sigmoidal) form of the function  $f$  is known. However, as shown in the Appendix, these priors can be disentangled from the observer's behavioural data. Moreover, the complexity of the relationship between them ensures that it would be possible, in principle, for an observer to respond *convex* to all 10 'upright' (ie convex with light-from-above) figures in (Stone and Pascalis (2010)), and still have a strong light-from-

*below* prior. This complexity also forces us to perceive hollow faces with overhead lighting as convex faces with light from below (see Figure 4), despite a default assumption that light comes from above.

**Possible confounds.** Previous studies have either been run in darkness (eg Adams et al. (2004); Stone et al. (2009)) or have explicitly manipulated the lighting direction (eg Mamassian and Landy (2001); Jenkin et al. (2004)). In this study, the natural lighting was always overhead, and was therefore in the same direction as the light-from-above prior. This confound seems to imply that we cannot be certain whether the children developed a stronger tendency to perceive shapes as illuminated above because their prior changed with age or because they became better at using lighting direction cues provided by the natural lighting present. However, in an experiment similar to that reported here Thomas et al. (2010) 51 children were tested at different ages using non-overhead (eye-level) lighting. Thomas’ data have a regression slope of +4.6% per year (estimated here from the regression line in their Figure 3). This compares to +4.2% per year (ie 0.84/20 items per year) found in the present study. The striking similarity in regression slopes in these two studies suggests that the natural lighting conditions used here had a negligible impact on the slope of the regression line, although it is logically possible that the physical lighting direction in the room did affect the height of the regression line.

One caveat applies to the results presented here. The observed increase in the probability of a ‘correct’ response could be caused either by an increase in the light-from-above prior  $\rho$  (as shown in Figure 2), or by a decrease in the decision-noise parameter  $\sigma_L$  (or by both). Alternatively, children may have access to high quality sensory data, and to a fully formed adult prior and likelihood, but their ability to integrate these by performing inference (which is known to be NP-hard (Paul and Michael (1993))) improves throughout childhood. However, the simple experimental design used here does not provide data that could be used, even in principle, to discriminate between these alternatives.

**Calibrating the Decision Noise Parameter Value:** If we assume a fixed value for the decision-noise parameter  $\sigma_L$  then we can justify the value of  $\sigma_L = 1.0$  used here, as follows. Define  $a$  to be the age at which the LFA prior has a neutral value of  $\rho = 0.5$ , and assume that the convexity prior is 0.5 (an assumption justified on the grounds that the estimated convexity prior is about 0.5 for all ages, almost irrespective of the value of  $\sigma$ ). If the LFA prior has a neutral value of  $\rho = 0.5$  at age  $a$  then it follows that the proportion of correct responses is also exactly 0.5 at age  $a$  (according to Equation 18), so we can estimate  $a$  from the re-



gression line drawn through the proportion of correct responses at different ages (ie from the regression line for combined picture and symbol stimuli in Figure 3a of Stone and Pascalis (2010)). The regression analysis reported in Stone and Pascalis (2010) predicts that the age at which the proportion of correct responses is 0.5 is 2.8 years, and this is therefore the age at which the LFA prior estimate is also 0.5. So, we know the age  $a = 2.8$  at which the LFA prior is equal to 0.5; what we do *not* know is  $\sigma_L$ . But, if the type of model used here is correct then there exists one value of  $\sigma_L$  which forces the regression line to pass through 0.5 at the known age  $a = 2.8$ . If we again use the combined picture and symbol stimuli then, as we vary putative values of  $\sigma_L$ , this changes the slope and intercept of the regression line for the LFA prior  $\rho$  (because each data point being fitted by the regression line is determined by the psychometric function defined by equation Equation 18). Numerical experiments indicate that the value of  $\sigma_L = 1.0$  used here provides a reasonable fit inasmuch as the combined picture and symbol stimuli yields a regression line (not shown) which predicts the LFA prior has a value of 0.5 at an age of 2.5 years. Thus, setting  $\sigma_L$  to a value that respects this approximate equality may be considered as a form of calibration for the decision-noise parameter. However, as the regression line for the LFA prior is fitted to data points which depend non-linearly (ie via Equation 18) on the value of  $\sigma_L$ , it is not obvious how to compute error bars for its value.

**Relation to Previous Work:** Evidence which suggests that infants have non-neutral priors by the age of 6 months was presented in (Granrud et al. (1985)) and (Tsuruhara et al. (2009)), and data re-analysed in (Stone and Pascalis (2010)) from (Yonas et al. (1979)) suggests an age of 9 months. If we equate the age at which children's performance is at chance (in terms of proportion correct) with the age at which the LFA prior is 0.5 (see previous paragraph) then the regression analyses reported in (Stone and Pascalis (2010)) predict a neutral LFA prior (0.5) at 1.7 years (for picture stimuli), and 3.8 years for symbol stimuli. Aside from differences in the stimuli used, the other difference between these studies is that, here, an 'in/out' verbal response was required for each single stimulus, whereas a 2AFC procedure was used in the three other studies cited above. Also, the estimates made here are based on *linear* regression lines from older children's data (4.6-10.8 years), rather than direct testing of infants, and it is possible that the actual change in non-linear below the age of 4 years.

**The Problem With Bayes' Theorem.** There is a great deal of debate regarding the status of Bayes' theorem, a debate which remains unresolved, but which perhaps should have been made largely redundant by the

work of Jaynes (2003) and Cox (1961). Jaynes states that much of the heat of this debate depends on ideological arguments regarding the interpretation of probability, and that there is actually little to debate if the mathematical infrastructure and scientific utility of different methods are compared. Specifically, in setting up rules for calculating with probabilities, we would insist that the result of such calculations should tally with our everyday experience of the physical world. That is, if we insist that probabilities must be combined with each other in accordance with certain common sense principles then Cox's consistency theorems lead to a unique set of rules; a set of rules which includes Bayes' theorem. In essence, this means that any other rules violate fundamental notions of rationality or consistency. So Bayes' theorem is not just a convenient rule which happens to work, it is part of a set of rules which are logically implied by our insistence that such rules should yield results which conform to the behaviour of the physical world.

## **Place Figure 4 about here.**

**The Problem With Priors.** When it is claimed that the prior distribution for lighting direction has been estimated, what does this mean in practice? In this study, there are two variable parameters, lighting direction ( $\theta$ ) and the convexity/concavity ( $c$ ) of a number of different shapes. As there is no reason to expect these two parameters to be correlated in the physical world, it makes sense to assume that they are independent. This means that the joint prior distribution  $p(\theta, c)$  can be represented as the product of two one-dimensional prior distributions, so that  $p(\theta, c) = p(\theta)p(c)$ . This independence not only makes the problem tractable, it also ensures that each of the estimated one-dimensional prior distributions is just a re-scaled marginal distribution of a multivariate (ie 2D) prior distribution.

However, the assumption of independence between parameters may not be justified in general, because some variables are correlated with others in the physical world, as shown in Figure 4. For example, supposing shape and lighting direction were not independent. For example, if Figure 1a denoted by  $c = 1$  is presented at multiple orientations in an experiment, and we estimated the prior distribution for lighting direction then what have we found? From Figure 4, it is apparent that we have *not* found the prior for lighting direction in general. Instead, we have found a cross-section of the 2D joint distribution  $p(\theta, c)$  at  $c = 1$ , which can be *considered* to be the prior for lighting direction, but is *actually associated with only a single shape*.

This argument also applies to the abbreviated, two-valued form of the lighting direction (above/below)

and convexity (convex/concave) prior distributions estimated here. Indeed, this argument may explain why the lighting-direction priors for symbols and pictures appear to be different (ie because they are different). Even when considered over both types of stimuli, we may not have estimated the prior  $p(\theta)$  for lighting (for example), but a kind of average prior for lighting associated with symbols and pictures. Only if the range of stimuli were sufficiently broad could we claim to have estimated the prior for lighting direction, because this broad range effectively marginalises the joint prior distribution  $p(c, \theta)$  over shape  $c$  leaving the marginal (prior) distribution  $p(\theta)$ . Given that the picture stimuli involved a broad range of pictures, this suggests that the ‘prior’ for pictorial stimuli is a better approximation to the lighting direction prior  $p(\theta)$  than the ‘prior’ associated with the relatively homogeneous set of symbol stimuli. The point is that, ideally, the estimated prior distribution is the marginal distribution of a two-dimensional joint distribution  $p(\theta, c)$ , as in Figure 4, but in practice, the estimated prior is more likely to be a cross-section of of the 2D prior  $p(\theta, c)$ . If this cross-section is considered for a face-like shape then the resultant lighting-direction ‘prior’ could well be uniform, because even hollow faces (eg from a mask) are perceived as convex, despite the fact that this forces the perceived lighting to come from below (Frisby and Stone (2010), p309). Even if this prior is not exactly uniform in adults, results presented here suggest that it is likely to be uniform in infants, which predicts that infants perceive hollow faces as convex, irrespective of the lighting direction.

More generally, the joint distribution  $p(\theta, c)$  is itself a marginal distribution of a high-dimensional prior with axes that include parameters such as shape, illuminance spectrum, multiple light sources, colour, texture, slant, size, depth, spatial/temporal frequency, and stereo disparity. If the parameters associated with these axes are independent then the high-dimensional prior can be represented as a set of 1D marginal prior distributions, each of which can be learned through exposure to the physical world. But if these parameters are not independent then this presents a serious problem for the brain, because the problem of learning the structure of the entire high-dimensional prior scales exponentially with the number of parameters (dimensions). Learning an adequate representation of such a prior could take many life-times observing the physical world, so it is likely to be hard-wired into the brain at birth, at least in an approximate form.

In practice, it is likely that subsets of physical parameters are mutually independent of other subsets. This effectively decomposes the problem of learning a single high-dimensional prior into the problem of learning many subsets of low-dimensional priors. These may be partly innate, but may also require fine-

tuning during childhood.

Naturally, there are considerations beyond the dimensionality of a given prior. For example, adult herring gulls have a red patch on their beaks, and herring gull chicks peck at this to elicit regurgitation of food from the adults (Tinbergen (1951)). Indeed, these chicks peck at almost anything that is red because this is critical to their survival. In such cases, there is little question about the innateness of the chick's ability to peck at red patches. Similarly, if humans were born into a world where milk could be obtained from convex, but not concave, objects where the shape was defined only by shading information, then it is likely that our ability to discriminate between convex and convex shapes would be almost perfect from birth.

## 5 Conclusion

The research reported here embodies the first attempt to track the developmental of Bayesian priors (rather than task performance) for shape convexity and lighting direction in children. As with most first attempts, it is far from perfect, but it is intended to represent a significant step towards a formal account of how Bayesian priors develop throughout childhood.

The results reported in Part I of this paper (Stone and Pascalis (2010)), suggest that children interpret ambiguous shading information from embossed symbol and naturalistic picture stimuli in an increasingly adult-like manner as they grow older, and the results reported here suggest that this age-related change in performance is contingent on an underlying change in children's prior assumption about lighting direction, but not on their assumption about shape convexity.

These results do not rule out the possibility that children have an innate, but weak, predisposition for interpreting stimuli as if light comes from above. However, whether or not children have a nascent light-from-above prior at birth, these results suggest that the prior for lighting direction is gradually learned throughout childhood as an empirical fact about the nature of the physical world.

Even though light has rarely come from below during the 540 million years since eyes first evolved, and even though almost all objects are essentially convex, these facts do not seem to have been hard-wired into the human brain in the form of strong innate light-from-above and convexity Bayesian priors.

**Acknowledgments:** I would like to thank Steve Isard and John Porrill for useful discussions, Charles Fox

for comments on an early draft of this paper, two anonymous referees for their detailed and insightful comments, and Sebastian and Teleri Stone for help with pilot experiments.

## References

- Adams, W., Graf, E., and Ernst, M. (2004). Experience can change the light-from-above prior. *Nature Neuroscience*, 7(10):1057–1058.
- Berbaum, K., Bever, T., and Chung, C. S. (1983). Light source position in the perception of object shape. *Perception*, 12(4):411–416.
- Berbaum, K., Bever, T., and Chung, C. S. (1984). Extending the perception of shape from known to unknown shading. *Perception*, 13(4):479–488.
- Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios, and of intaglios into cameos, with an account of other analogous phenomena. *Edinburgh Journal of Science*, 4:99–108.
- Cox, R. (1961). *The Algebra of Probable Inference*. Johns Hopkins University Press, Baltimore.
- Frisby, J. and Stone, J. (2010). *Seeing: The computational approach to biological vision*. MIT Press.
- Gold, J. and Shadlen, M. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cog Sci*, 5:10–16.
- Granrud, C., Yonas, A., and Opland, E. (1985). Infants sensitivity to the depth cue of shading. *Perception and Psychophysics*, 37:415–419.
- Green, D. and Swets, J. (1966). *Signal detection theory and psychophysics*. John Wiley and Sons.
- Hershberger, W. (1970). Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below. *J Comp Physiol Psychol*, 73:407–411.
- Hess, E. (1950). Development of the chick's responses to light and shade cues of depth. *Journal of Comparative and Physiological Psychology*, 43(2):112–122.
- Jaynes, E. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge.
- Jenkin, H., Jenkin, M., Dyde, R., and Harris, L. (2004). Shape-from-shading depends on both shading and perceived orientation. *Perception*, 33:1453–1461.

- Kleffner, D. and Ramachandran, V. (1992). On the perception of shape from shading. *Perception and Psychophysics*, 52(1):18–36.
- Liu, B. and Todd, J. (2004). Perceptual biases in the interpretation of 3D shape from shading. *Vision research*, 44(18):2135–2145.
- Mamassian, P. and Landy, M. (2001). Interaction of visual prior constraints. *Vision Research*, 41:2653–2668.
- Metzger, W. (1936 (reprinted 2009)). *Laws of Seeing*. MIT Press.
- Paul, D. and Michael, L. (1993). Approximating probabilistic inference in bayesian belief networks is np-hard. *Artificial Intelligence*, 60(1):141153.
- Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317:314–319.
- Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society*, 2:37–42.
- Stone, J., Kerrigan, I., and Porrill, J. (2009). Where is the light? Bayesian perceptual priors for lighting direction. *Proceedings Royal Society London (B)*, 276:1797–1804.
- Stone, J. and Pascalis, O. (2010). Footprints sticking out of the sand (part i): Children’s perception of naturalistic and embossed symbol stimuli. *Perception*.
- Sun, J. and Perona, J. (1998). Where is the sun? *Nature Neuroscience*, 1(3):183–184.
- Thomas, R., Nardini, M., and Mareschal, D. (2010). Interactions between light-from-above and convexity priors in visual development. *Journal of Vision*.
- Tinbergen, N. (1951). *The Study of Instinct*. Oxford University Press.
- Tsuruhara, A., Sawada, T., Kanazawa, S., Yamaguchi, M. K., and Yonas, A. (2009). Infant’s ability to form a common representation of an object’s shape from different pictorial depth cues: A transfer-across-cues study. *Infant Behavior and Development*, 32(4):468 – 475.

Von Fieandt, K. (1949). Das Phanomenologische Problem von Licht und Schatten (the phenomenon of light and shadow). *Acta Psychologica*.

Woodworth, R. and Schlosberg, H. (1954). *Experimental Psychology (3rd Edition)*. Methuen, London.

Yonas, A., Kuskowski, M., and Sternfels, S. (1979). The role of frames of reference in the development of responsiveness to shading information. *Child Development*, 50:495–500.



## Appendix: Estimating Priors

The shape information in each stimulus  $\mathbf{x}$  is a function of two parameters, the direction  $\theta$  of the light source, and the degree  $c$  of convexity/concavity of the 3D shape depicted in the image. When presented with an image  $\mathbf{x}$  of a shaded object, the perceived shape  $\hat{c}$  and lighting direction  $\hat{\theta}$  depends upon the observer's prior expectations about shape and lighting direction as represented in the observer's prior probability density function (pdf)  $p(c, \theta)$ , and on the likelihood function  $p(\mathbf{x}|c, \theta)$ , which is the probability of the image given a shape  $c$  and lighting direction  $\theta$ .

The perceived shape  $\hat{c}$  is a function of the posterior pdf  $p(c, \theta|\mathbf{x})$ , and is usually assumed to correspond to the pair of values of  $(c, \theta)$  that maximises  $p(c, \theta|\mathbf{x})$ . Using Bayes' rule we can express the posterior in terms of the prior and the likelihood,

$$p(c, \theta|\mathbf{x}) = p(\mathbf{x}|\theta, c)p(c, \theta)/p(\mathbf{x}), \quad (1)$$

where  $p(\mathbf{x})$  is treated as a constant and we assume  $p(\mathbf{x}) = 1$ , which allows us to omit it from now on. It is usually assumed that the observer has an accurate estimate of the likelihood function.

**Loss function:** The loss function defines how costly errors in perceived shape are to an observer. We assume that the visual system minimises average loss  $E$ , where this average is taken over values of  $c$  and  $\theta$

$$E = \int_c \int_\theta D(\hat{c}, c) p(c, \theta|\mathbf{x}) d\theta dc = 1/p(\mathbf{x}) \int_c \int_\theta D(\hat{c}, c) p(\mathbf{x}|c, \theta) p(c, \theta) d\theta dc, \quad (2)$$

where  $D(\hat{c}, c)$  is a loss function. Here we assume a zero-one loss function, which adopts the value  $D(\hat{c}, c) = 0$  for correctly classified stimuli, and  $D(\hat{c}, c) = -1$  for incorrectly classified stimuli. If the observer makes choices that minimise  $E$  then the stimulus  $\mathbf{x}$  is perceived as the shape  $\hat{c}$  as if it is lit from direction  $\hat{\theta}$ , where (it can be shown that) the values  $c = \hat{c}$  and  $\theta = \hat{\theta}$  correspond to the maximum value of the joint posterior probability density function (pdf)  $p(c, \theta|\mathbf{x})$ .

It is important to note that the quantities referred to above are properties of an observer, rather than properties of the physical world. In other words, the objective is to estimate the prior distribution of an observer with respect to lighting direction and convexity, but not to estimate the prior distribution of lighting

direction in the physical world.

In practice, each stimulus  $\mathbf{x}$  used here is consistent with only two (opposite) lighting directions, which we define as light-from-above (LFA)  $\theta_a$  and light-from-below (LFB)  $\theta_b = \theta_a + 180^\circ$ , each of which is consistent with only one shape, either convex  $c_0$  or concave  $c_1$  (depending on the stimulus considered). These two classes of shapes/lighting directions will be referred to as convex/LFA and concave/LFB (eg for the images in Figure 1), and as convex/LFB and concave/LFA (eg for inverted versions of the images in Figure 1). In order to use a concrete example, we consider the case of a stimulus  $\mathbf{x}_a$  which is perceived as convex if light is assumed to come from above (ie convex/LFA), and concave if light is assumed to come from below (ie concave/LFB).

In order to minimise the number of mis-classified stimuli, signal detection theory (Green and Swets (1966)) dictates that, given a stimulus  $\mathbf{x}_a$ , the observer should respond *convex* if the posterior probability density that the shape is convex/LFA is greater than the posterior probability density that the shape is concave/LFB, or equivalently, if the posterior ratio

$$R = \frac{p(c_0, \theta_a | \mathbf{x}_a)}{p(c_1, \theta_b | \mathbf{x}_a)}, \quad (3)$$

is greater than one, and *concave* otherwise. Applying Bayes' rule (Equation (1)) to the numerator and denominator of Equation (3) yields

$$R = \frac{p(\mathbf{x}_a | c_0, \theta_a)}{p(\mathbf{x}_a | c_1, \theta_b)} \times \frac{p(c_0, \theta_a)}{p(c_1, \theta_b)} \quad (4)$$

$$= \frac{p(c_0, \theta_a)}{p(c_1, \theta_b)}. \quad (5)$$

Crucially, the stimulus  $\mathbf{x}_a$  is *equally* consistent with two physical scenarios, namely, convex/LFA and concave/LFB. Assuming that the observer's likelihood function reflects this equality, the observer's two likelihood values in Equation (4) are equal (ie  $p(\mathbf{x}_a | c_0, \theta_a) = p(\mathbf{x}_a | c_1, \theta_b)$ ). This ensures that the ratio of posterior values in Equation (3) is equal to a ratio of prior values, as shown in Equation (5).

**The Key Step:** Equations (3-5) are key to understanding the method. The reason why the likelihood values are equal in Equation (4) is that a single stimulus  $\mathbf{x}_a$  is equally likely to have been generated (as in Figure 1, for example) either by a convex shape with light from above (convex/LFA), or by a concave shape with

light from below (concave/LFB). Crucially, if the likelihood values cancel (as they do for the stimuli used here) then the ratio of posterior values is the same as the ratio of prior values. This is important because we can use the measured observer responses to estimate the ratio of posterior values, which we now know to be equal to the ratio of prior values, and this can be used to estimate the prior values.

**The Posterior Pdf**  $p(c|\mathbf{x}_a)$ : Having shown how equal likelihood values lead to a ratio of prior values, we now back up a little in order to derive an expression for the posterior pdf  $p(c|\mathbf{x}_a)$ . In principle, the observer's posterior pdf for shape  $p(c|\mathbf{x}_a)$  is obtained by marginalising over  $\theta$ :

$$p(c|\mathbf{x}_a) = \frac{1}{p(\mathbf{x}_a)} \int p(\mathbf{x}_a|c, \theta) p(c, \theta) d\theta. \quad (6)$$

However, as already noted, each stimulus  $\mathbf{x}_a$  is consistent with only two (opposite) lighting directions,  $\theta_a$  and  $\theta_b = \theta_a + 180^\circ$ . This implies that  $p(\mathbf{x}_a|c, \theta)$  is non-zero only at  $p(\mathbf{x}_a|c_0, \theta_a)$  and at  $p(\mathbf{x}_a|c_1, \theta_b)$ , which allows us to model the likelihood function  $p(\mathbf{x}_a|c, \theta)$  as a pair of delta functions

$$p(\mathbf{x}_a|c_0, \theta) = \delta(\theta - \theta_a) \quad \text{and} \quad p(\mathbf{x}_a|c_1, \theta) = \delta(\theta - \theta_b). \quad (7)$$

Substituting Equation (7) in the integral of Equation (6) for  $c = c_0$  yields a cross-section  $p(c_0, \theta)$  of the joint prior at  $c = c_0$ , which is the posterior

$$p(c_0|\mathbf{x}_a) = \frac{1}{p(\mathbf{x}_a)} \int \delta(\theta - \theta_a) p(c_0, \theta) d\theta, \quad (8)$$

$$= p(c_0, \theta_a)/p(\mathbf{x}_a). \quad (9)$$

A similar line of reasoning for  $c = c_1$  yields the posterior

$$p(c_1|\mathbf{x}_a) = p(c_1, \theta_b)/p(\mathbf{x}_a). \quad (10)$$

Now define

$$\bar{L}_0 = \log p(c_0|\mathbf{x}_a), \quad \text{and} \quad \bar{L}_1 = \log p(c_1|\mathbf{x}_a). \quad (11)$$

As before (but now using these newly defined log-posteriors), the observer should respond *convex* if  $\bar{L}_0 > \bar{L}_1$  and *concave* otherwise. Equivalently, the observer should respond *convex* only if the log posterior ratio

$$\bar{L} = \log \frac{p(c_0|\mathbf{x})}{p(c_1|\mathbf{x})} \quad (12)$$

$$= \bar{L}_0 - \bar{L}_1, \quad (13)$$

is greater than zero. Substituting the posteriors of Equations (9) and (10) into Equation (12) yields the log *prior* ratio

$$\bar{L} = \log \frac{p(c_0, \theta_a)}{p(c_1, \theta_b)}. \quad (14)$$

If the observer assumes that the stimulus shape and the lighting direction are mutually independent then the joint prior distribution  $p(c, \theta)$  factorises,  $p(c, \theta) = p(c)p(\theta)$ , so that Equation (14) can be re-written as

$$\bar{L} = \log \frac{p(c_0) p(\theta_a)}{p(c_1) p(\theta_b)}, \quad (15)$$

where (for example)  $p(\theta_a)$  is the value of the prior for the LFA lighting direction  $\theta_a$ , and  $p(c_0)$  is the value of the prior for convexity  $c_0$ .

Regardless of the lighting direction, each observer perceives each stimulus as either convex  $c_0$  or concave  $c_1$ , and responds accordingly. Thus, together,  $p(c_0)$  and  $p(c_1)$  is a pair of co-determined observer-specific scalar priors, such that  $p(c_1) + p(c_0) = 1$ . For the two lighting directions (above/below) considered here, the same logic implies that  $p(\theta_a) + p(\theta_b) = 1$ . We call the scalar  $p(c_0)$  the *convexity prior*, and the scalar  $p(\theta_a)$  the *light-from-above (LFA) prior*, for a given observer. Notice that we have discretized the lighting directions into  $N = 2$  values:  $\theta_a$  and  $\theta_b$ . This effectively models the entire prior distribution for lighting direction as the pair  $p(\theta) = (p(\theta_a), p(\theta_b))$ .

**Modelling Noisy Responses:** As human decision making is noisy, we assume that the process which compares the measured value of the log probability density  $\bar{L}_0$  with  $\bar{L}_1$  is subject to noise, in a manner broadly consistent with the signal detection theory (Green and Swets (1966)). Specifically, the measured values of

$\bar{L}_0$  and  $\bar{L}_1$  are assumed to be, respectively,

$$L_0 = \bar{L}_0 + \eta_0 \quad \text{and} \quad L_1 = \bar{L}_1 + \eta_1, \quad (16)$$

where  $\eta_0$  and  $\eta_1$  are Gaussian noise terms, both with zero mean and standard deviation  $\sigma$ . Therefore the distribution of  $L_0$  values is Gaussian with mean  $\bar{L}_0$  and the distribution of  $L_1$  values is Gaussian with mean  $\bar{L}_1$ , both with standard deviation  $\sigma$ . This implies that  $L = L_0 - L_1$  is also Gaussian with mean  $\bar{L} = \bar{L}_0 - \bar{L}_1$  and variance  $\sigma_L^2 = 2\sigma^2$ , where  $\sigma_L$  is a decision-noise parameter.

The use of log-ratios can be justified because the neural encoding of log probabilities has a simple neural interpretation (Gold and Shadlen (2001)), such that a difference in spike rate between two neurons constitutes a decision variable that is proportional to the log ratio  $L$ .

Given these assumptions, the probability  $P(c_0|\mathbf{x})$  that the observer perceives the shape  $c_0$  in the stimulus  $\mathbf{x}_a$  is given by the probability that  $L_0 > L_1$ , which is described by the cumulative density function of a Gaussian

$$P(c_0|\mathbf{x}_a) = P(L_0 > L_1) \quad (17)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\bar{L}/\sigma_L} e^{-\eta^2/2} d\eta \quad (18)$$

$$= \Phi(\bar{L}/\sigma_L), \quad (19)$$

where the ratio  $\bar{L}/\sigma_L$  can be interpreted as a z-value with zero mean and unit variance<sup>3</sup>. Given a measured value for the probability  $P(c_0|\mathbf{x})$  of a convex response, we can obtain  $\bar{L}/\sigma_L$  from the inverse function

$$\bar{L}/\sigma_L = \Phi^{-1}(P(c_0|\mathbf{x})). \quad (20)$$

Note that the decision-noise parameter  $\sigma_L$  is measured in units of log-probability density, and (for the delta functions assumed here) is independent of the sensory-noise parameter  $\sigma_\theta$  conventionally associated with signal detection theory.

---

<sup>3</sup>A slightly different definition of  $\sigma_L$  was used in (Stone et al. (2009)), which we denote as  $\sigma_{2009}$ , and in (Stone et al. (2009)),  $\sigma_{2009} = 2$ . The value of  $\sigma_L = 1$  used here is equivalent to a value of  $\sigma_{2009} = 4.34$ .

**The Method in a Nutshell:** At this point, we have enough information to summarise the key steps used to estimate the observer’s prior pdf.

1. If we know the functional relationship between the measured probability  $P(c_0|\mathbf{x}_a)$  that the observer responds *convex* and the log ratio of posterior probability densities  $\bar{L}$  (which we do, via Equation (20)) then we can use  $P(c_0|\mathbf{x}_a)$  to estimate  $\bar{L}/\sigma_L$ .
2. For each stimulus used here, we have shown that the log ratio of posterior probability densities is equal to the log ratio of prior probability densities.
3. Under mild assumptions, we can use the log ratio of prior probability densities to derive estimates for the prior probability densities (ie not just their ratios).

Having outlined the key steps, we give details of these below.

**Maximum Likelihood Estimation:** For a stimulus  $\mathbf{x}_a$  that can be perceived as convex only if light is assumed to come from above, the probability of a convex response is defined as  $q_a = P(c_0|\mathbf{x}_a)$ . In contrast, the inverted version of this stimulus  $\mathbf{x}_b$  can be perceived as convex only if light is assumed to come from below, and the probability of a convex response is defined as  $q_b = P(c_0|\mathbf{x}_b)$ . For the full set of  $n_a = 10$  symbol and picture stimuli, which can be perceived only as either convex/LFA ( $c = c_0, \theta = \theta_a$ ) or concave/LFB ( $c = c_1, \theta = \theta_b$ ), the probability of a convex response  $q_a$  defines (via Equation (20)) the log prior ratio

$$\bar{L}_a/\sigma_L = \frac{1}{\sigma_L} \log \frac{p(c_0) p(\theta_a)}{p(c_1) p(\theta_b)}. \quad (21)$$

Similarly, if the  $n_b = 10$  pictures in Figure 1 are *inverted* then they can be perceived only as either convex/LFB ( $c = c_0, \theta = \theta_b$ ) or concave/LFA ( $c = c_1, \theta = \theta_a$ ), and the probability of a convex response  $q_b$  defines the log prior ratio

$$\bar{L}_b/\sigma_L = \frac{1}{\sigma_L} \log \frac{p(c_0) p(\theta_b)}{p(c_1) p(\theta_a)}. \quad (22)$$

For the  $n_a = 10$  stimuli used here (see two examples in Figure 1), the measured number  $m_a$  of *convex*

( $r = 0$ ) responses defines a likelihood function for the probability  $q_a$  of a convex response as

$$p(m_a|q_a) = q_a^{m_a}(1 - q_a)^{n_a - m_a}, \quad (23)$$

where  $n_a - m_a$  is the number of of *concave* responses, and the constant binomial coefficient  $C_{n_a, m_a}$  has been omitted because it will not affect our final estimate of  $q_a$ . Notice that, given  $n_a = 10$  stimuli which can be perceived as either convex/LFA or concave/LFB, the number  $m_a$  of *convex* responses can be considered to be the number of ‘votes’ for the LFA lighting direction  $\theta_a$  and the convex shape  $c_0$ ; and the number of *concave* responses  $n_a - m_a$  can be considered to be the number of ‘votes’ for the LFB lighting direction  $\theta_b$  and the concave shape  $c_1$ .

Similarly, for the  $n_b = 10$  stimuli (as in an inverted version of Figure 1), the number  $m_b$  of *convex* ( $r = 0$ ) responses defines the likelihood function

$$p(m_b|q_b) = q_b^{m_b}(1 - q_b)^{n_b - m_b}. \quad (24)$$

The values  $\hat{q}_a$  and  $\hat{q}_b$  for  $q_a$  and  $q_b$  that maximise Equations (23) and (24), respectively, are maximum likelihood estimates (MLE), and are given by the proportion of convex responses

$$\hat{q}_a = m_a/n_a \quad \text{and} \quad \hat{q}_b = m_b/n_b. \quad (25)$$

At this point, we can either find estimates of the two priors using an easy but opaque method, or using a transparent but hard method. The easy opaque method consists of using a numerical search for prior values that maximise the product of the likelihoods defined in Equations (23) and (24). The hard transparent method is described below, and is intended to provide some insight into how the priors are estimated. Both methods provide the same maximum likelihood estimates (MLE) of the prior values.

For any measured value of  $q_a$  and  $q_b$ , we can obtain  $\bar{L}_a/\sigma_L$  and  $\bar{L}_b/\sigma_L$  via Equation (20) (recall that  $q_a = P(c_0|\mathbf{x}_a)$ ). So, thus far, we have two known terms,  $\bar{L}_a/\sigma_L$  and  $\bar{L}_b/\sigma_L$ . Equations (21) and (22) can

then be used to express their difference in terms of the log ratio of prior values at  $\theta_a$  and  $\theta_b$

$$\frac{\bar{L}_a - \bar{L}_b}{\sigma_L} = \frac{1}{\sigma_L} \log \frac{p(c_0) p(\theta_a)}{p(c_1) p(\theta_b)} - \frac{1}{\sigma_L} \log \frac{p(c_0) p(\theta_b)}{p(c_1) p(\theta_a)} \quad (26)$$

$$= \frac{2}{\sigma_L} \log \frac{p(\theta_a)}{p(\theta_b)} \quad (27)$$

$$= \Delta \bar{L}, \quad (28)$$

where  $\Delta \bar{L}$  is defined for convenience, and its value is known because it is the difference between our two known terms ( $\bar{L}_a/\sigma_L$  and  $\bar{L}_b/\sigma_L$ ). If we assume that  $p(\theta_a) + p(\theta_b) = 1$  then we can substitute  $p(\theta_b) = 1 - p(\theta_a)$  in Equation (27), which can be re-arranged to obtain an estimate of the LFA prior as a logistic function of  $\Delta L$

$$\hat{p}(\theta_a) = 1/(1 + e^{-\sigma_L \Delta \bar{L}/2}). \quad (29)$$

Having obtained  $\hat{p}(\theta_a)$ , this is used to find an estimate  $\hat{p}(c_0)$  the convexity prior, as follows. First, define

$$R_c = p(c_0)/p(c_1) \quad (30)$$

$$= p(c_0)/(1 - p(c_0)) \quad (31)$$

$$R_\theta = p(\theta_a)/p(\theta_b) \quad (32)$$

$$= p(\theta_a)/(1 - p(\theta_a)), \quad (33)$$

so that Equation (21) can be written as

$$\bar{L}_a/\sigma_L = \frac{1}{\sigma_L} \log \frac{R_c}{R_\theta}. \quad (34)$$

Multiplying both sides by  $\sigma_L$ , and re-arranging yields

$$R_c = R_\theta e^{\bar{L}_a}. \quad (35)$$

Re-arranging Equation (31) yields

$$p(c_0) = 1/(1 + 1/R_c), \quad (36)$$



and substituting Equation (35) in (36) yields

$$\hat{p}(c_0) = 1/(1 + 1/(R_\theta e^{\bar{L}_a})), \quad (37)$$

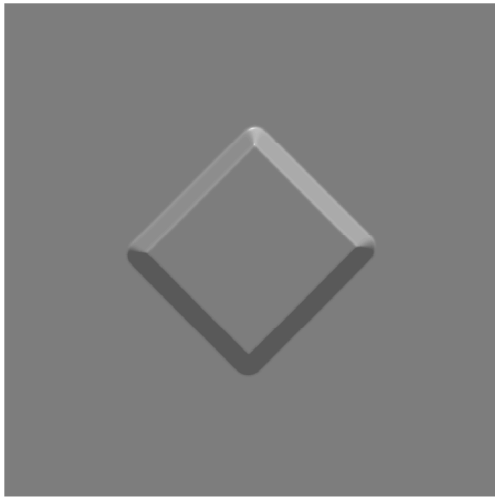
where  $R_\theta$  is known from Equation (33), and the LFB prior estimate is obtained as  $\hat{p}(c_1) = 1 - \hat{p}(c_0)$ .

In summary, the maximum likelihood estimates (MLE)  $\hat{q}_a$  and  $\hat{q}_b$  are found from the observer's responses using Equations (25) and (25). These can then be used in the steps defined in Equations (27-29) to obtain an estimate  $\hat{p}(\theta_a)$  of the LFA prior, which can be used in Equations (30-36) to obtain an estimate  $\hat{p}(c_0)$  of the convexity prior. Thus, the LFA prior and the convexity prior can be estimated from the observer's responses. The functional invariance property of maximum likelihood estimates ensures that these estimates, which are derived from the MLE  $\hat{q}_a$ , are also maximum likelihood estimates.

**Estimating Priors for Simulated Observers.** The method for estimating priors was tested using data from a set of 20 simulated observers, a number which is similar to the number of children in each age group used here. All 20 simulated observers had the same LFA prior  $p(\theta_a) = 0.8$  and the same convexity prior  $p(c_0) = 0.6$ , with  $\sigma_L = 1$ . Each simulated observer gave stochastic binary 'responses' in accordance with Equation (18) for  $n_a = 5$  stimuli and  $n_b = 5$  'inverted' versions of those stimuli (eg for the picture stimuli). The responses from each simulated observer were then used to obtain an estimate of the LFA and convexity prior for that simulated observer, using the method described above for actual observers. Considered over all 20 simulated observers, the mean LFA prior was estimated as  $\bar{p}(\theta_a) = 0.823$  (sem=0.0162), and the mean convexity prior was estimated as  $\bar{p}(c_0) = 0.612$  (sem=0.0235), neither of which is significantly different from the known values of these priors (0.8 and 0.6, respectively). The results of these numerical experiments suggest that the method provides accurate estimates of the LFA and convexity priors.

**Setting the value of  $\sigma_L$ .** If the value of the decision-noise parameter  $\sigma_L$  is set too low then numerical problems arise which prevent accurate estimates of priors. For example, setting  $\sigma_L = 1$ , if the LFA prior  $p(\theta_a) = 0.8$  and the convexity prior  $p(c_0) = 0.6$  then the upper limit of the integral in Equation (18) is  $L/\sigma_L = 1.79$  which yields  $P(c_0|\mathbf{x}) = 0.96$  (where  $L/\sigma_L$  is a z-score). If  $\sigma_L$  is set much lower than unity then  $P(c_0|\mathbf{x})$  values are forced towards zero or one for most values of the convexity and LFA priors. Numerical experiments indicate that setting  $\sigma_L \leq 0.5$  leads to under-estimates of the prior values, and that, in order to recover prior values over a reasonable range, it is necessary to set  $\sigma_L \geq 0.75$ , and we use  $\sigma_L = 1$

in this paper. Overall, using data from the 171 children, increasing the value of  $\sigma_L$  increases the rate at which the LFA prior increases with age, but leaves the rate at which the convexity prior increases with age fairly constant at about zero. (A slightly different definition of  $\sigma_L$  was used in (Stone et al. (2009)), which we denote as  $\sigma_{2009}$ , and in (Stone et al. (2009)),  $\sigma_{2009} = 2$ . The value of  $\sigma_L = 1$  used here is equivalent to a value of  $\sigma_{2009} = 4.34$ ).



(a)



(b)

Figure 1: **(a) Example of symbol stimulus, and, (b) picture stimulus.** Each stimulus was presented as here, and upside-down, and the observer indicated whether the shape appeared to be convex ('out') or concave ('in'). Both stimuli here should be perceived as convex if the observer assumes light comes from above (ie from the top of the page). Conversely, turning the page upside-down forces the observer to perceive these stimuli as concave if light is assumed to come from above. Note that these are 2 samples of the 10 actual stimuli used, and that the embossed appearance of (a) was represented by a flat (ie non-embossed) stimulus. The full set of stimuli can be seen in (Stone and Pascalis (2010)). Figure b is reproduced with permission from Jan Sevcik.

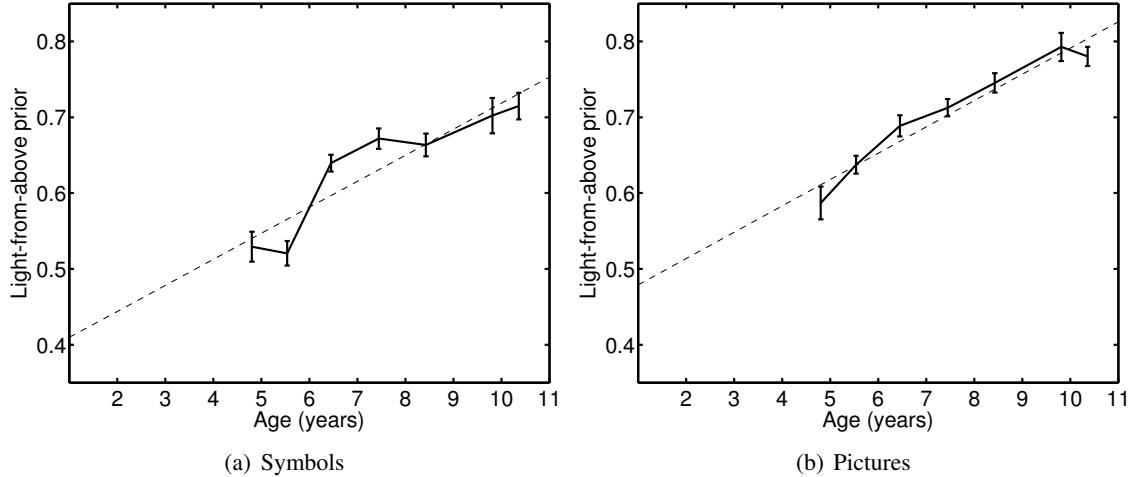


Figure 2: **Changes in the light-from-above (LFA) prior with age for (a) symbol stimuli and (b) picture stimuli.**

(a) For symbols, the mean LFA prior is  $\rho_{sym} = 0.629$  (sem= 0.013). A regression analysis of priors  $\rho_{sym}$  against age yielded  $\rho_{sym} = 0.034 \times \text{age} + 0.375$  ( $R^2 = 0.839$ ,  $F = 35.9$ ,  $p = 0.004$ ), which predicts a neutral lighting direction prior (0.5) at 3.63 years.

(b) For pictures, the mean LFA prior is  $\rho_{pic} = 0.700$  (sem= 0.012). A regression analysis of priors  $\rho_{pic}$  against age yielded  $\rho_{pic} = 0.035 \times \text{age} + 0.444$  ( $R^2 = 0.947$ ,  $F = 89.1$ ,  $p < 0.001$ ), which predicts a neutral lighting direction prior (0.5) at 1.61 years.

In each graph, the dashed line is a fitted regression line, and error bars denote standard errors.

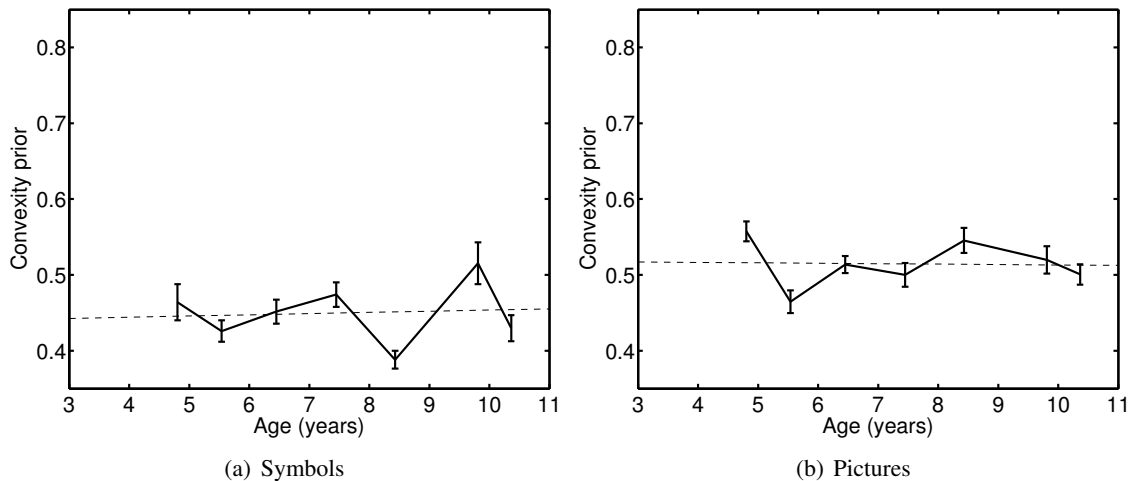


Figure 3: **Changes in convexity prior with age for (a) symbol stimuli and (b) picture stimuli.**

(a) Symbol convexity priors. The mean convexity prior is 0.446 (sem= 0.013). A regression of convexity prior  $\gamma_{sym}$  against age yielded  $\gamma_{sym} = 0.002 \times \text{age} + 0.438$  ( $R^2 = 0.007$ ,  $F = 0.034$ ,  $p = 0.862$ ).

(b) Picture convexity priors. The mean convexity prior is 0.509 (sem= 0.011). A regression of convexity prior  $\gamma_{pic}$  against age yielded  $\gamma_{pic} = -0.001 \times \text{age} + 0.519$  ( $R^2 = 0.002$ ,  $F = 0.008$ ,  $p = 0.934$ ).

In each graph, the dashed line is a fitted regression line, and error bars denote standard errors.

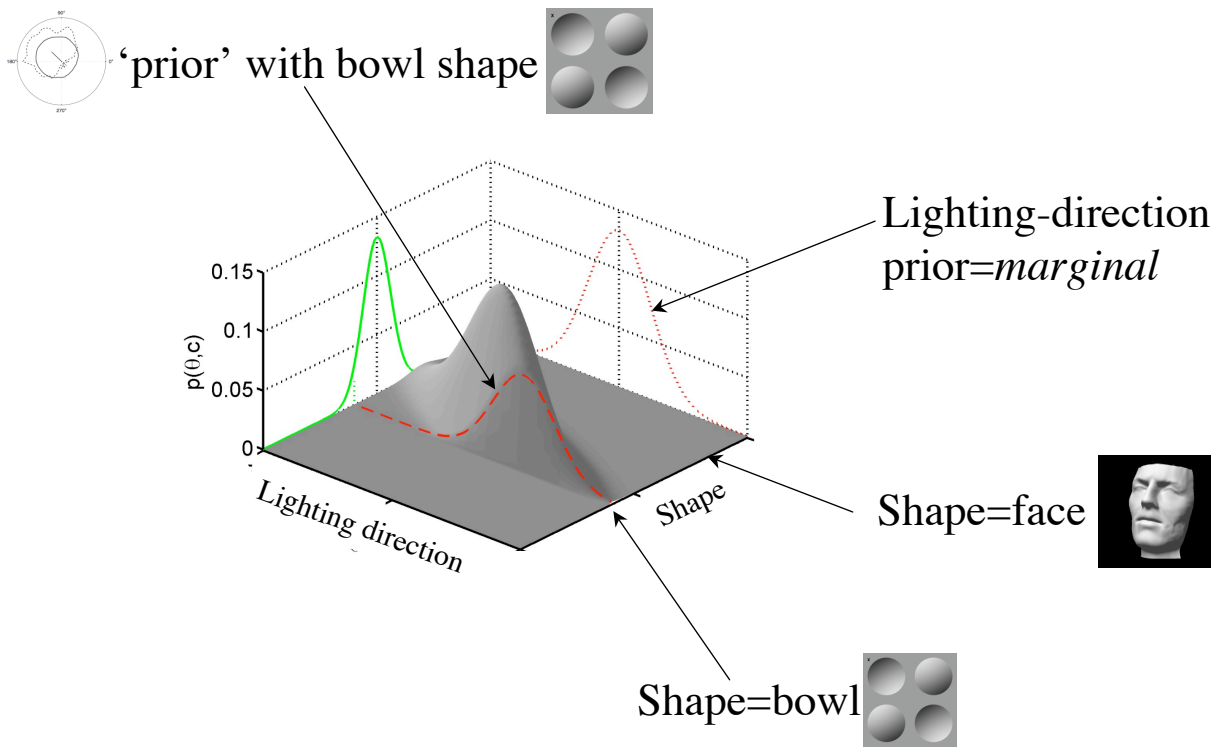


Figure 4: **The problem with priors.** A hypothetical multivariate (2D) observer’s prior for shape  $c$  and lighting direction  $\theta$  defines a surface with a height given by  $p(\theta, c)$ . The prior for shape is the marginal distribution  $p(c)$  (the solid curve on the left hand wall), and the prior for lighting direction is the marginal distribution  $p(\theta)$  (the dotted curve on the right hand wall). For example, the shape parameter  $c = 3$  could specify the face-like stimulus shown. If we used the bowl-like stimuli shown in an experiment (eg with shape parameter  $c = 1$ ) then the measured ‘prior’ for lighting direction corresponds to a cross-section of the multivariate prior  $p(\theta, c)$  at  $c = 1$ , given by the dashed curve. If the two parameters  $c$  and  $\theta$  are independent then the measured ‘prior’ would be a scaled version of the observer’s true prior. However, if  $c$  and  $\theta$  are correlated (as shown here) then the measured ‘prior’ is the dashed curve, which is a scaled and *shifted* version the observer’s true prior, where the amount of shift depends on the particular shape  $c$  used to measure the observer’s prior.