**Bayesian models for scene-level Adelson illusion**

Chengxu Zhuang (chengxuz@stanford.edu)

Department of Psychology, Stanford

Andrea Shulman (ashulman@stanford.edu)

Department of Psychology, Stanford

# Abstract

Optical illusions provide insight into the mechanics of perception. Using WebPPL, a probabilistic programming language, we have created models of optical illusions and have generated new backwards illusions. These models can help us understand how the human mind perceives visual images. To simulate the cognition perspective of an illusion, we first break the Adelson illusion into two separate pictures: one consisting only of contrast and another that considers a three-dimensional object and its shadow. The first part of the illusion is better expressed as a simpler, Simultaneous Contrast Illusion. The model for this illusion is used as a base algorithm for producing the model for the Adelson illusion. With reflectance, illumination, and luminance as the main parameters, we used human perception of color and lighting to force the mind to misconstrue the actual pixel value of the color. Our algorithm generated new optical illusions in which two different shades of gray looked more alike than two of the same shades of gray. This research shows how easy it is to trick the human brain with simple shading and object positioning.

**Keywords:** illusion; Bayes; probabilistic programming;

reflectance; illuminance; luminance; WebPPL

# Introduction

The visual system is one of the most astonishing systems in the human body with its brilliant ability to process and interpret information from input (such as physical light) to a contextual representation of the surrounding world. However, it can sometimes be too intricate and leave room for confusion by carefully designed illusions. The underlying neural mechanism for those illusions is still poorly understood [2]. One possible way of solving this is to model the human observers inferring the perceived colors using Bayesian models [2]. Bayesian models have been successfully used in modeling both many higher-order cognitive functions [3] or lower-order sensory systems [4]. Through this project, we attempt to model illusions through Bayesian analysis and run these models in reverse to create new illusions. Ultimately, we want to understand more about the intricacy of the mind when it interprets these optical illusions.

The checkershadow illusion (Figure 1) [5], designed by Edward H. Adelson of MIT [1], is famous for having a shadow on a checkerboard that tricks the brain into thinking two of the same shaded squares are very differently colored. Illusions like Adelson illusions (A-illusion) are useful for testing how the brain perceives illumination and shading and uses them to build models to understand the environment. By using Bayesian modeling of those illusions, we could not only shed light on the underlying mechanism of brain integrating related information, but also provide an effective and generative model for making better and more illusions and to possibly explain related neural dynamics.

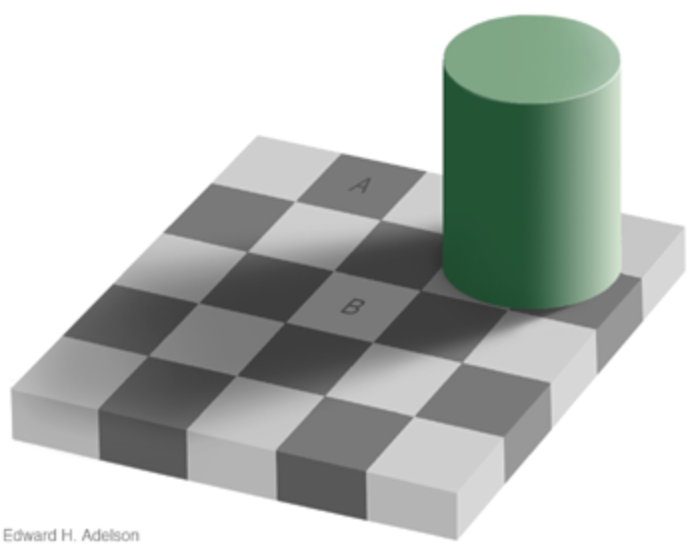


Figure 1. Adelson illusion [5]

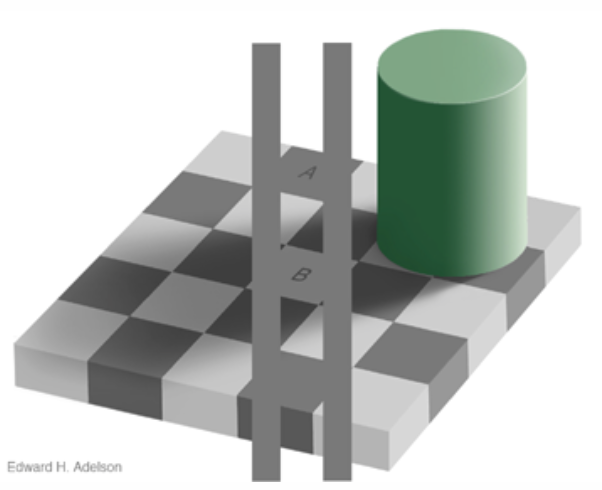


Figure 2. Explanation to Adelson illusion [5]

As indicated by Adelson [5], A-illusion is actually a result of at least two different levels of inference by human observers. First, observers actively interpolate the darker surface as a shadow of the 3D object and other areas as brighter areas without shadows. Secondly, observers combine the information of the neighbors to infer the area of interest and its illuminance (amount of light cast on the surface) and reflectance (amount of reflected light).

The second level can be extracted and investigated independently. Therefore, we chose to evaluate a simpler illusion first: The Simultaneous Contrast Illusion (SC-illusion). The Simultaneous Contrast Illusion (see Figure 3) also uses context to trick our minds [1], where the varying background luminescence makes the inscribed shapes appear darker or lighter than the identical other. Because of the absence of three dimensional objects (and thus shadows), it is expected that human observers simply use the local information for inference as they would in the second level of A-illusion. Understanding the basics, the SC-illusion, will make it easier to understand the more complex A-illusion.



Figure 3. SC-illusion

After showing the modeling and results for SC-illusion, we return to A-illusion, by modeling the inferences made by human observers as a hierarchical Bayesian model, where the higher level is inferencing the shadow map for the entire picture and the lower level is a modified version of Bayesian models for SC-illusion with shadow information better utilized.

With both the illusions modeled, the models can be run in reverse thus generating new backwards illusions. The possible result of this could be illusions where two different shades of gray appear to be the same shade of gray to the human eye. Illusion research could implement our resulting pictures and understand more about human cognition.

# Related Work

There have been various research papers explaining illusions in different levels using different methods.

Brainard et al. [4] has tried to use Bayesian inference models to explain those similar illusions. They proposed the illumination, reflectance, and luminance models where they used local pixel values to infer illumination. Our model for SC-illusion is similar to theirs. But they did not use hierarchical Bayesian models for A-illusion.

It is mostly believed that the underlying neural mechanism might be related to the properties of ON- and OFF-center ganglion cells [7]. However, this is still not entirely understood and needs to be more deeply researched [4].

After being proposed by Adelson [1], these illusions have been hot topics for vision research for many years [2,3,4]. There are also proposed similar illusions under the name ‘color constancy’ [8].

## Methods

## Human observer model outline

Following, we briefly describe the Bayesian inference models that attempt to mimic brains that are interpreting these illusions. In our hypothesis, we suppose brains are actively updating two variables for an area-of-interest: illumination () and reflectance (), where denotes an area.

Both and should be a spectrum across different wavelength of light. But in our models, we only consider grayscale illusions. Therefore, we just consider those two variables as scalars where represents how much light is cast on area and represents how much of the light is reflected by area . Higher is saying that more light is reflected and therefore area is lighter, when lower means that area is darker.

Human observers have some prior for both and . In our models, those prior distributions are usually Gaussian distributions. Then they would also have a formula for luminance () (by physical rules) and likelihood function for actually observed pixel values (). We simply take as a noisy observation of , i.e., .

Using those observations, they could infer the posterior distributions of (). In this paper, we are especially interested in because it is the perceived color and what needs to be altered to for an illusion to trick the brain.

# Extension to basic model for SC- and A-illusion

Before explaining our results, we first introduce the basic models we currently have for human observers viewing SC-illusion and A-illusion.

SC-illusion is comparatively easier, as there is no 3D object and human observers do not need to infer anything about whether there are shadows in this area. For simplification, we refer the two center squares as L-AOI and R-AOI (L/R for left/right, and AOI for area of interest). SC-illusion shows that human observers would perceive the color (reflectance in our model) for L/R-AOI as different, while they actually have the same pixel value (luminance in our model). Our hypothesis is that this is due to the different backgrounds of those two AOIs with the assumption that human observers are inferring both the illumination and reflectance based on the local luminance observed. And we model the influence of neighbors by introducing only one assumption: the illumination should be spatially continuous. The intuition here is that without the existence of shadows, human observers would interpret the whole area under the same lighting condition thus assuming areas nearby should also have similar illumination. This could be a result of not using high-resolution inference map about illumination towards this illusion, possibly due to larger receptive fields and small number of neurons responsible for representing illumination in human brains.

But for A-illusion, the existence of the 3D object is indicating that some of the surface is under shadows and should have lower illumination. Thus, we need to explicitly model a different mechanism of inferring the illumination with or without shadows, as well as the influence of different backgrounds. This requires two additional steps beyond completing a model of the SC-illusion. The first step is to determine which areas are under shadows of 3D objects (which need to be modeled). Then, we would have two different models of inferring illumination from luminance observed (for shadow areas and areas without shadows). Therefore, we are effectively modeling A-illusion using a hierarchical Bayesian model.

## Results

## Modeling SC-illusion

The picture in Figure 1 has the size of pixels (with the AOI in the size of pixels), but we group the pixels into “superpixels” with each superpixel having size of pixels and indexing them using , where ranges from 0 to 5 and ranges from 0 to 2. We do this for simplification and because human observers tend to interpret space of the same color as a large continuous region rather than many pixels of the same color. To continue, the L-AOI is in the position of and R-AOI is at . As indicated above, each superpixel has its own reflectance (), illumination () and luminance (). And the pixel value observed is . If we assume that each superpixel is independent from each other, we could have a model as below ( means has a prior of a Gaussian distribution centered at with standard deviation of , is the collection of all parameters, and represents the collection of all ):

To introduce the influence of backgrounds, we build the continuous assumption for illumination into .

The reason why we build this assumption into prior of is that we believe this continuous assumption is due to the biological limit such as receptive field and number of neurons used to represent illumination map.

Thus the the existence of spatial continuity for the inference of illumination could be modelled as below:

Here, is the set of indexes for nearby superpixels of :

To apply this continuity, we first draw from independently. And after combining the observations of , we could get a posterior distribution of for the observation of spatial continuity. We then use this posterior distribution as .

Combining those priors and observations, we could then compute the posterior distribution of and , which is the perceived colors of L-AOI and R-AOI respectively.

## Experiment results for SC-illusion

We implement the models using WebPPL [9], with , , , and . The pixel values used during inference are 1 for dark regions (neighbors of R-AOI), 3 for both L-AOI and R-AOI, and 5 for brighter regions (neighbors of L-AOI). See Figure 3 for the computed posterior distribution of .



Figure 4. Distribution of , with a mean of -0.52

The result shown has already explained the illusion observed by human observers. But for further explorations of the model, we use the same mechanism to build a reverse SC-illusion (rSC-illusion) which now have different pixel values for L/R-AOIs but by utilizing the influence of backgrounds, make the human observers believe that they see the same colors. We could examine whether our model is modelling human observers correctly by simply looking at rSC-illusion generated by the model and judge how similar do two AOIs look like.

The implementation to build this rSC-illusion is straight-forward. Instead of providing the observations for L/R-AOIs of their pixel values, we add an observation that their reflectance is the same:

And then we infer the posterior distribution for the luminance and use the inferred value to draw a new picture. The pixel values used in the image generating (pixel values in 0-255) are calculated from pixel values used in inference by . See Figure 5 for rSC-illusion below:



Figure 5. rSC-illusion, with pixel value of 159 (L-AOI) and 127 (R-AOI), other pixel values are the same as them in SC-illusion

Both the distribution and rSC-illusion shown above have shown that our model could explain SC-illusion correctly. In the next section we will introduce our model and results for A-illusion.

## Modeling A-illusion

As described above, the theoretic models for human observers viewing A-illusion should be a hierarchical Bayesian model.

In higher level inference, models infer a binary variable for each location representing whether this location is under shadow or not and another global categorical variable for the direction of light.

We first introduce how we infer and then how influences .

The existence of shadows in A-illusion is due to the existence of 3D object (a cylinder in our situation) and that it seems this cylinder is blocking some light. Especially, we could infer the direction of light based on the side of this cylinder, as around half of the side is darker than another half. The darker region would be interpreted as the result of the light not hitting that area thus provide information about direction of light.



Figure 6. Illustration of cutting of cylinder side

Specifically, we first cut the side of cylinder into 8 vertical areas where in A-illusion (indexed from 0 to 7), where only the first 4 areas are visible to us, as shown in Figure 6. Different direction of light could result in different brightness of these eight areas. The direction of light should be a continuous scalar describing the degree of angles between light and vertical axis. But for simplification, we assume that direction of light in our problem could only take eight different values, indexed from 0 to 7. While , we set area 3~6 to be bright. And while , we set area 4~7 to be bright. The bright areas of other directions of light could be set in the similar way. We set the pixel values observed of area 0~3 in A-illusion to be [1, 1, 1, 3]. The pixel values of these eight areas are represented by , range from 0 to 7. And see below for the models to infer .

Combine observations of , , , , we could get posterior distribution of . We then take this posterior distribution as the prior of () in further inference.

To infer , besides , we also need the position of the object. Before doing that, we need to group the pixels in A-illusion to “superpixels” similarly to what we have done for SC-illusion. The image is abstracted to superpixels and every superpixel is indexed using , where ranges from 0 to 4 and ranges from 0 to 4. And the 3D object is occupying 4 superpixels: , , , and . The pixel value of area A and B in inference is set to be 3. The neighbors of A are set to be 6 and that of B are set to be 1. We set 9 superpixels to be in darker area. For brighter area, black region in checkerboard would have pixel value of 3 while white region would have pixel value of 6. And for darker area, black region would be 1 and white region would be 3. The 9 superpixels set to be in darker area are: , , , , , , , , and . To describe the meaning of in this space, we set to represent the light having the same direction from to . And is light from to . Other directions could be similarly described.

After combining and object position, we could make some inferences about . And after having , we could sample based on that and then use similar model in SC-illusion for inferences about . The whole model is described as below:

For three prior functions used in the model, has been described above. And for , given , we could first divide all superpixels to 2 sets (, ) by judging whether the straight line between any object superpixels and this superpixel would be parallel to the direction of light. If yes, that superpixel would belong to , otherwise it would belong to . Therefore could be sampled from prior function described in detail as below:

And for prior function of , we first build two priors for the whole map () in shadow or not (the original ). If in shadow, we just replace the prior of during the construction of from to . And if , we sample from , otherwise, we sample from .

As we build two priors independently, illuminations of regions in would be independent from that of regions in , as we don’t have spatial continuity for regions across two sets.

Combining the model described above and all observations, we could then get our posterior distribution for A-illusion. We would describe them in details below.

## Experiment results for A-illusion

We also implement this model using WebPPL [9], with , , . The other parameters are the same as that in SC-illusion.

We could get the distribution of as shown in Figure 7 below:

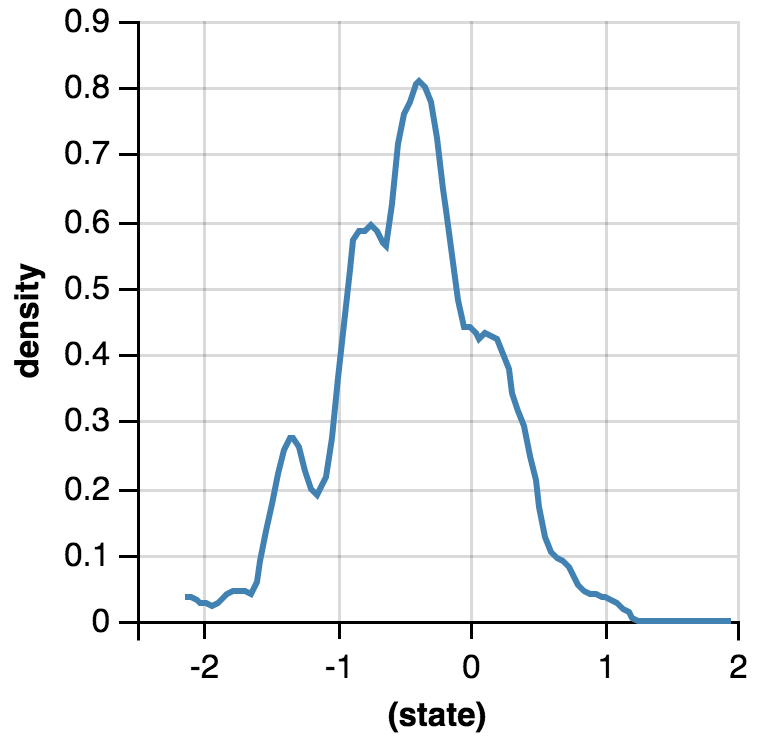


Figure 7. Distribution of , with a mean of -0.41.

As shown in Figure 7, the result already shows that is smaller than , which means we could already explain A-illusion.

But similar to SC-illusion, we build a reverse illusion by adding the observation of reflectance of area A and B being the same and inferring the illuminance of them. The pixel value shown in the picture is computed by . And the result is shown in Figure 8.

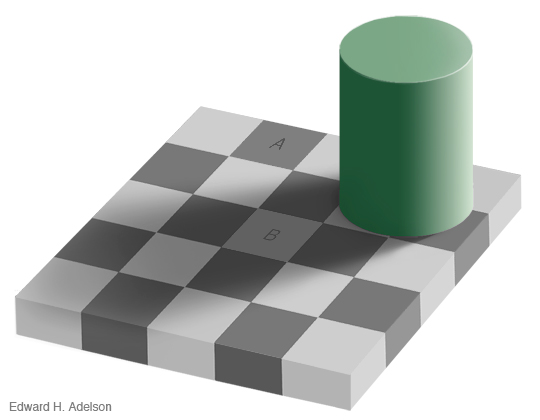


Figure 8. rA-illusion, with pixel value of 127 in A and 97 in B, other pixel values are the same as them in A-illusion

And the results shown above illustrates that our model has successfully explained this illusion and the reverse illusion also proves that our model is similar to what human observer is actually using for A-illusion. But there are still many improvements that could be made. We would discuss them in the next section.

## Discussion

Our paper builds models for both SC-illusion and A-illusion. We could successfully explain the illusion experienced by human observers by only introducing very simple and reasonable assumptions. We also run our models of these illusions backwards to generate what we call reverse illusions. And the reverse illusions we generate show that our model could model human observers well that it could tricky human observers by knowing what they will see and infer.

There are many improvements we could make to our current models.

In our current models, we need to first infer a complicated posterior distribution with continuous assumption building in and then use that as a prior function. Instead of doing this, we could build a hierarchical Bayesian model to make continuous assumption as a result of this.

And in our model for A-illusion, we are inferring the posterior distribution of direction of light first and then use that as a prior distribution in the joint inference about both shadow variables and illumination. One possible improvement is to include the inference of direction of light into the joint inference. We tried this but found that there would be too many weak evidence about direction of light in the joint distribution to make the evidence provided by object pixel values ignore by the final inference, which is not what we want.

Different level of analysis could provide different insights about the system of interest. In our work, we use Bayesian analysis to explain how human observers are tricked by optical illusions. Although Bayesian model is usually thought to be irrelevant to lower level, such as underlying neural mechanism for this. We are still interested in how our model could be implemented by neural circuits. One hypothesis we have proposed is that the central assumption in our model, the continuous assumption might be due to large receptive field of neurons representing illuminations, as larger receptive field means lower precision to determine exact illumination for a small area rather than a big area. But it is still not clear how the inferring part of our model could be implemented using neural circuits.

The reverse illusion we generate could further be used to investigate the correct parameters in the models. Additionally, we could crowdsource the human observers’ responses to reverse-illusions of different parameters. And then we could use the data collected to build models to explain relationships between parameters in our models to what is happening inside brains of human observers.

During our investigations, we found that after knowing the mechanism of the illusion and even the information that there was an illusion, we would lose our ability to be tricked by this illusion or implicitly correct the illusion we see. We call this phenomenon post-illusion. Inside the framework of our models, we could potentially explain this by adjusting the continuity for illumination. After knowing the illusion, we might increase the resolution of inference map about illumination, which means the spatial continuity of illumination will be decreased, and therefore the illusion will also be decreased.

In conclusion, our models could already explain how human observers are tricked by SC-illusions and A-illusions. But they still have many potentials to include other related phenomenon or increase the reasonability of the models. We expect that these could be finished in further work about this. And these could help we better understand and explain the inferring algorithms used by human observers towards those illusions.

# Acknowledgments

We would like to express our gratitude to Michael Henry Tessler for his patient guidance and Professor Noah Goodman for his wonderful lessons.

# References

[1] Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science, 262*(5142), 2042-2044.

[2] Allred, S. R., & Brainard, D. H. (2013). A Bayesian model of lightness perception that incorporates spatial variation in the illumination. *Journal of vision, 13*(7), 18-18.

[3] Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in cognitive sciences, 10*(7), 287-291.

[4] Brainard, D., & Gazzaniga, M. S. (2009). Bayesian approaches to color vision. *The visual neurosciences, 4*.

[5] Adelson, E. H. (1995). *Checkershadow Illusion.* Retrieved from <http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html>.

[6] *Bayesian inference*. (2016, November 18). In Wikipedia, The Free Encyclopedia. Retrieved 18:42, November 18, 2016, from <https://en.wikipedia.org/w/index.php?title=Bayesian_inference&oldid=750266512>

[7] Famiglietti, E. V., & Kolb, H. (1976). Structural basis for ON-and OFF-center responses in retinal ganglion cells. *Science, 194*(4261), 193-195.

[8] Maloney, L. T., & Wandell, B. A. (1986). Color constancy: a method for recovering surface spectral reflectance. *JOSA A*, *3*(1), 29-33.

[9] N. D. Goodman and A. Stuhlmüller (electronic). The Design and Implementation of Probabilistic Programming Languages. Retrieved from http://dippl.org.