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 Bayesian 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 involves 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. This model takes what people interpret from an image, and changes the aspects people do not consciously consider. 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 that Bayesian models can explain how human observers process an optical illusion.

Keywords: illusion; Bayes; probabilistic programming; WebPPL; reflectance; illuminance; luminance

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. For example, contemplate Figure 1 and 2: blocks A and B appear differently colored but they are the same. The underlying neural mechanism for those illusions is still poorly understood (Allred & Brainard, 2013). One possible way of solving this is to model the human observers inference of the perceived colors using Bayesian models (Allred & Brainard, 2013). Bayesian models have been successfully used in modeling both many higher-order cognitive functions and lower-order sensory systems (Chater, Tenenbaum, & Yuille, 2006; Brainard & Gazzaniga, 2009). 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), designed by Edward H. Adelson of MIT, 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 (Adelson, 1993; Adelson, 1995). 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 to make better and more illusions and to possibly explain related neural dynamics.

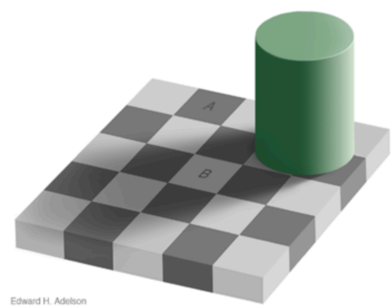

Figure 1. Adelson illusion (Adelson, 1995)

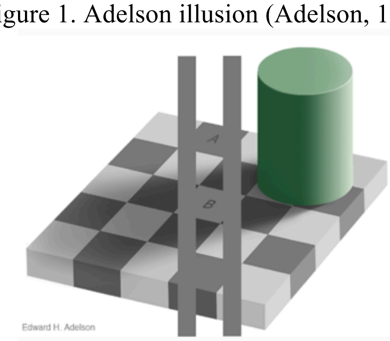

Figure 2. Explanation to Adelson illusion (Adelson, 1995)

As indicated by Adelson, A-illusion is actually a result of at least two different levels of inference by human observers (Adelson, 1995). 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 (Adelson, 1993), 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 and Gazzaniga have tried to use Bayesian inference models to explain those similar illusions (2009). 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 (Famiglietti & Kolb, 1976). However, this is still not entirely understood and needs to be more deeply researched (Brainard & Gazzaniga, 2009).

After being proposed by Adelson, these illusions have been hot topics for vision research for many years (Adelson, 1993; Allred & Brainard, 2013; Chater, Tenenbaum, & Yuille, 2006; Brainard, D., & Gazzaniga, 2009). There are also proposed similar illusions under the name ‘color constancy’ (Maloney & Wandell, 1986).

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 (I_{pos}) and reflectance (r_{pos}), where pos denotes an area.

Both I_{pos} and r_{pos} should be a spectrum across different wavelength of light. But in our models, we only study grayscale illusions. Therefore, we only deem those two variables as scalars where I_{pos} represents how much light is cast on area pos and r_{pos} represents how much of the light gets reflected by area pos. Higher r_{pos} signifies that more light is reflected and, therefore, area pos is lighter, while lower r_{pos} indicates that area pos is darker.

Human observers have some prior for both I_{pos} and r_{pos}. In our models, those prior distributions are usually Gaussian distributions. The product of these distributions is the luminance ($lu_{pos} = I_{pos}r_{pos}$) (by physical rules) and the likelihood function for actual observed pixel values is $P\left(p_{pos} \middle| I_{pos}, r_{pos}\right)$. We take p_{pos} as a noisy observation of lu_{pos}, i.e., $P(p_{pos}) \sim N\left(lu_{pos}, \sigma_{lu}\right)$.

These observations help us infer the posterior distributions of I_{pos}, r_{pos} ($P\left(I_{pos}, r_{pos} \middle| p_{pos}\right)$). In this paper, we are especially interested in r_{pos} because the perceived color is what needs to be altered for an illusion to trick the mind.

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, human observers infer both the illumination and reflectance based on the observed local luminance. 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 conditions, thus assuming areas nearby should also have similar illumination. This could be a result of not using a high-resolution inference map of illumination on this illusion or could be 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. The next step is to have two separate models of inferring illumination from observed luminance – one for both areas with and without shadows. This allows us to effectively model A-illusion using a hierarchical Bayesian model.

Results

Modeling SC-illusion

The picture in Figure 1 has the size of 180×90 pixels (with the AOI in the size of 30×30pixels), but we group the pixels into 6×3 "superpixels" with each superpixel having size of 30×30 pixels and indexing them using (i, j), where i ranges from 0 to 5 and j ranges from 0 to 2. We do this for simplification and because, intuitively, 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 $(1,1)$ and R-AOI is at $(4,1)$. As indicated above, each superpixel has its own reflectance ($r_{i,j}$), illumination ($I_{i,j}$) and luminance ($lu_{i,j}$). And the pixel value observed is $p_{i,j}$. If we assume that each superpixel is independent, we could have a model, $r_{i,j}\sim N(\mu_R, \sigma_R)$. This means $r_{i,j}$ has a prior of a Gaussian distribution centered at μ_R with standard deviation of σ_R. w is the collection of all parameters, and $\boldsymbol{I}$ represents the collection of all $I_{i,j}$.

$$I_{i,j}\sim\mathrm{Prior}(\boldsymbol{I}),$$
$$r_{i,j}\sim N(\mu_R, \sigma_R),$$
$$lu_{i,j} = r_{i,j}\times I_{i,j}$$
$$P\left(p_{i,j}\middle|w\right)\sim N\left(lu_{i,j}, \sigma_{lu}\right)$$

To introduce the influence of backgrounds, we build the continuous assumption for illumination into $\mathrm{Prior}(\boldsymbol{I})$. We also do this because we believe this assumption is due to the biological limit. Such limits include receptive fields and number of neurons used to represent illumination map.

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

$$DI_{i,j} = I_{i,j} - \mathrm{mean}(\{I_{k,l}|(k,l) \in A_{i,j}\})$$
$$P(DI_{i,j})\sim N(0, \sigma_{DI})$$

Here, $A_{i,j}$ is the set of indexes for nearby superpixels of (i, j):

$$A_{i,j} = \{(k,l)|\mathrm{abs}(k-i) + \mathrm{abs}(l-j) = 1,$$
$$k \in [0,5], j \in [0,2]\}$$

To apply this continuity, we first draw $I_{i,j}$ from $N(\mu_I, \sigma_I)$ independently. After combining the observations of $P(DI_{i,j})\sim N(0, \sigma_{DI})$, we could get a posterior distribution of $\boldsymbol{I}$ for the observation of spatial continuity. This posterior distribution is then used as $\mathrm{Prior}(\boldsymbol{I})$.

Combining those priors and observations, we could then compute the posterior distributions of $r_{1,1}$ and $r_{4,1}$, which represent the perceived colors of L-AOI and R-AOI respectively.

Experiment results for SC-illusion

We implement the models using WebPPL (Goodman & Stuhlmüller, 2014), with $\mu_I = 2$, $\mu_R = 1$, $\sigma_{DI} = 0.3$, and $\sigma_I = \sigma_R = \sigma_\alpha = \sigma_l = 1$. 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 4 for the computed posterior distribution of $r_{1,1} - r_{4,1}$.

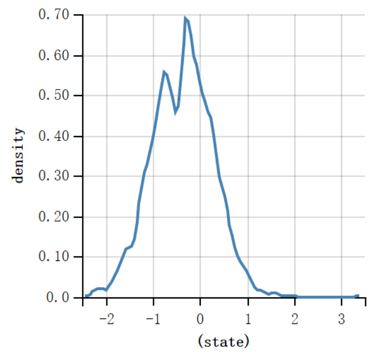

Figure 4. Distribution of $r_{1,1} - r_{4,1}$, with a mean of -0.36

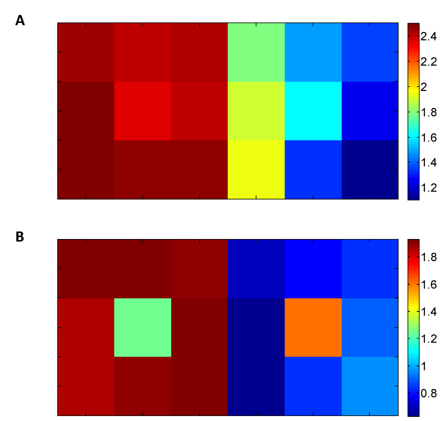

Figure 5. A. The map of expectations of inferred illuminations. B. The map of expectations of inferred reflectance.

The result shown has already explained the illusion observed by people. We also plotted the illumination and reflectance maps inferred for SC-illusion in Figure 5. From the illumination map, it is clear that we assume the two AOIs' illumination is similar to that of their neighbors. As the left part of the image increases in luminance (pixel value), the model guesses this part also has higher illumination. Similar inference is also made on the right part of SC-illusion. The illuminations are different for L/R-AOIs, but their luminances are the same. This then causes the reflectance of these two areas is to be inferred differently.

For further explorations of the model, we use the same mechanism to build a reverse SC-illusion (rSC-illusion). This new illusion now has 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 exhibiting human observers correctly by simply looking at rSC-illusion generated by the model and judging how similar the two AOIs look.

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

$$P(r_{1,1} - r_{4,1} | w) \sim N(0, \sigma_{DI})$$

Then, we infer the posterior distribution of the luminance and use the inferred value to draw a new picture. The pixel values used in the image generated (pixel values in 0-255) are calculated from pixel values used in inference by $30 \times p_{i,j} + 50$. See Figure 6 for rSC-illusion below:

Figure 6. rSC-illusion: pixel values of 159 (L-AOI) and 127 (R-AOI), other pixel values are equal to those in SC-illusion

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

Modeling A-illusion

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

In higher level inference, models infer a binary variable $s_{i,j}$ for each location that represents if the area is under a shadow. Another global categorical variable ld is be used for the direction of light. We first introduce how we infer ld and then how ld influences $s_{i,j}$.

The existence of shadows in A-illusion is due to the existence of a 3-D object (a cylinder in our situation) which seems to be blocking some light. We could infer the direction of light based on the coloring of each side of the cylinder: one half is light and the other half is dark. The darker region would be interpreted as the part not receiving light. This provides us with information about the direction of light.

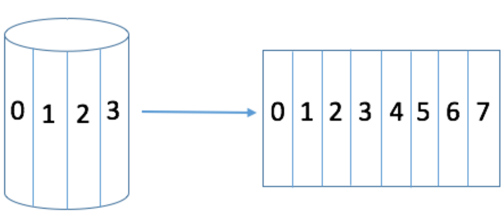

Figure 7. Illustration of cutting of cylinder side

Specifically, to determine the direction of the light, we first cut the side of cylinder into 8 vertical areas (indexed from 0 to 7). In A-illusion, only the first 4 areas are visible to us, as shown in Figure 7. Different light directions could result in varying brightness of these eight areas. The direction of light should be a continuous scalar describing the degree of angles between the light source and the vertical axis. However, for simplification, we assume direction of light can only take eight different values: 0 to 7. While $ld = 0$, we set area 3-6 to be bright. Also, while $ld = 1$, 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 ps_i, i range from 0 to 7. See below for the models to infer ld.

$$ld \sim \text{uniform}(0,1,2,3,4,5,6,7)$$

$$P(ps_i | ld) \sim N(E_{i,ld}, \sigma_{ld})$$

$$E_{i,ld} = \begin{cases} 1, & i \text{ in bright areas of } ld \\ 3, & \text{otherwise} \end{cases}$$

We combine observations of ps_0, ps_1, ps_2, ps_3 to get the posterior distribution of ld. We then take this posterior distribution as a prior of ld (Prior(ld)) for further inference.

To infer $s_{i,j}$, without ld, the position of the object is needed. However, first, we need to group the pixels in A-illusion into "superpixels" as we did for SC-illusion (see Figure 7). The image is reduced to a 5×5 map of superpixels and every superpixel is indexed using (i, j), where i ranges from 0 to 4 and j ranges from 0 to 4. The 3D object is occupying 4 superpixels: (3,3), (3,4), (4,3), and (4,4). The pixel value of area A and B 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 the darker area. For the brighter area, the darker squares in the checkerboard would have pixel values of 3 while lighter regions would have pixel values of 6. For the darker area, the darker regions would be 1 and lighter regions would be 3. The 9 superpixels set to be in the darker area are: (0,0), (0,1), (1,0), (1,1), (1,2), (2,1), (2,2), (2,3), and (3,2). To describe the meaning of ld in this space, we set $ld = 0$ to represent the light having the same direction from (0,0) to $(-1,-1)$. $ld = 1$ is light from (0,0) to $(-1,0)$. Other directions could be similarly described. See Figure 8.

After combining ld and object position, we could make some inferences about $s_{i,j}$. Then, we could sample $I_{i,j}$ and use a similar model in SC-illusion for inferences about $r_{i,j}$. The whole model is described below:

$$ld \sim \text{Prior}(ld),$$

$$s_{i,j} \sim \text{Prior}(\boldsymbol{s} | ld),$$

$$I_{i,j} \sim \text{Prior}(\boldsymbol{I} |\, s_{i,j}),$$

$$r_{i,j} \sim N(\mu_R, \sigma_R),$$

$$lu_{i,j} = r_{i,j} \times I_{i,j},$$

$$P(p_{i,j} | w) \sim N(lu_{i,j}, \sigma_{lu})$$

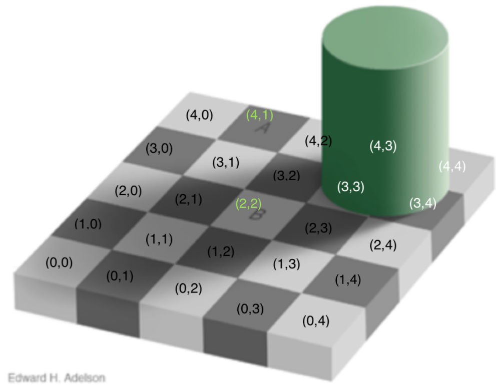

Figure 8. Superpixel indexes map

Out of the three prior functions used in the model, $\text{Prior}(ld)$ has been described above. For $\text{Prior}(\boldsymbol{s}|ld)$, we first divide all superpixels to 2 sets (Bs, Ss) by judging whether a straight line between any of the object superpixels and the superpixel in question would be parallel to the direction of light. If so, that superpixel would belong to Ss, otherwise it would belong to Bs. Therefore $s_{i,j}$ could be sampled from the prior function described in detail below:

$$s_{i,j} \sim \begin{cases} \text{Bernoulli}(sp_B), (i,j) \text{ belongs to } Bs \\ \text{Bernoulli}(sp_S), (i,j) \text{ belongs to } Ss \end{cases}$$

For the prior function of $\boldsymbol{I}$, we first build two priors for the whole map that expresses ($\text{Prior}_S(\boldsymbol{I})$) if an area is in a shadow or not (the original $\text{Prior}(\boldsymbol{I})$). If an area is in a shadow, we change the prior of $I_{i,j}$ during the construction of $\text{Prior}(\boldsymbol{I})$ from $N(\mu_I, \sigma_I)$ to $N(\mu_I - \Delta\mu_S, \sigma_I)$. If $s_{i,j} = 1$, we sample $I_{i,j}$ from $\text{Prior}_S(\boldsymbol{I})$, otherwise, we sample from $\text{Prior}(\boldsymbol{I})$.

As we build two priors independently, illuminations of regions in Ss are independent from illuminations of regions in Bs. This is due to not having 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.

Experiment results for A-illusion

This model uses WebPPL (Goodman & Stuhlmüller, 2014), with $\sigma_{ld} = 0.5$, $sp_B = 0.95$, $sp_S = 0.05$, and $\Delta\mu_S = 2$. Other parameters are the same as in SC-illusion. The distribution of $r_A - r_B$ is shown in Figure 9 below.

The prior and posterior distributions of direction of light are similar: both peaked at 0 with a probability at around 0.95. The posterior distribution of $s_{i,j}$ for the darker places in our superpixel map is about 0.99 for 1, while that for the brighter places in our map is about 0.03 for 1. These probabilities are different from sp_B and sp_S, respectively, which means our inference does change it to the right direction.

We plot our inference of illumination and reflectance map in Figure 10. From the two maps, we find all shadow areas are clearly in a lower illumination and other areas have higher illumination. The border between shadow and non-shadow is clear in our model, which suits our intuition that areas with parts of shadows and non-shadows should not have continuous illumination. The reflectance map also shows we have explained the illusion successfully and our model does have the power to explain the illusion on a scene-level.

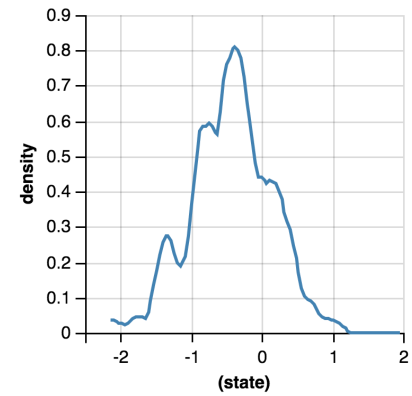

Figure 9. Distribution of $r_A - r_B$, with a mean of -0.41. This shows we can explain A-illusion.

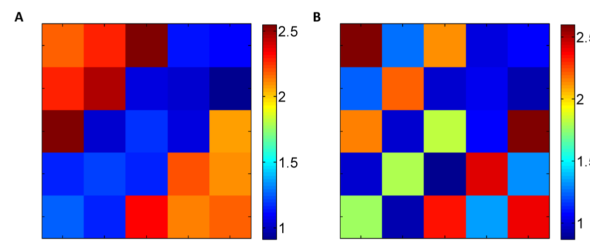

Figure 10. A. The map of expectations of inferred illuminations. B. The map of expectations of inferred reflectance.

Similar to SC-illusion, we build a reverse illusion by setting the reflectances of areas A and B equal, and inferring their illuminances. The pixel value shown in the picture is computed by $29 \times p_{i,j} + 33$.

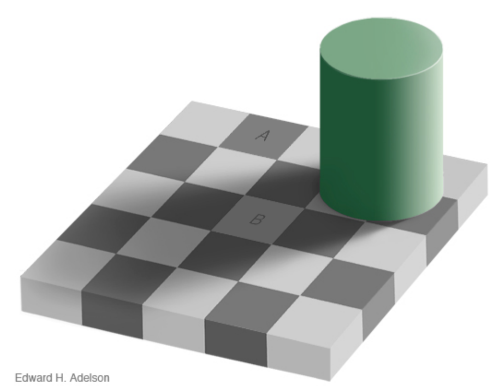

Figure 11. rA-illusion, with pixel value of 127 in A and 97 in B, other pixel values are the same as them in A-illusion

The results illustrate that our model has successfully explained this illusion. Also, our reverse illusion proves that our model is similar to what the human observer is actually using for A-illusion. However, there are still many opportunities for improvement.

Discussion

We have built models for both SC-illusion and A-illusion and now can explain more about how optical illusions are experienced by human observers. This only required introducing very simple and reasonable assumptions. We also ran our models of these illusions backwards to generate what we call reverse illusions. These new pictures illustrate that our model could mimic human observers well enough that it could trick the mind and identify what it will see and infer.

There are many improvements we could make to our current models. In our current models, we first need to infer a complicated posterior distribution with a continuous assumption built in, and then use that as a prior function. Instead of doing this, we could simply build a hierarchical Bayesian model to make the continuous assumption. Another improvement could be made in our model for A-illusion. Here, we are inferring the posterior distribution of direction of light first, then using that as a prior distribution in the joint inference of both shadow variables and illumination. A possible enhancement is to include the inference of direction of light in the joint inference.

For A-illusion, we use 25 superpixels to represent the original checkerboard. This representation could greatly reduce the number of variables needed, but also limit the power of representing the whole images. For example, the shadows in the image actually have a fuzzy border, which we could not properly convey. Adelson claimed that the fuzzy border would help the illusion trick the mind. We could use more superpixels to make modelling fuzzy border possible, but this would also require more inference.

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. The Bayesian model is on a different level from the underlying neural mechanism so they may not be related. However, we are still interested in how our model could be implemented by neural circuits.

The central assumption in our model, the continuous assumption, might be due to the large receptive field of neurons representing illuminations. Large receptive field means there is lower precision which makes it harder to determine the exact illumination for a small area. However, this is only one hypothesis, and it is still not clear how the inference part of our model could be implemented using neural circuits.

Furthermore, the reverse illusion we generated could be used to investigate the parameters humans correctly describe. Additionally, crowdsourcing human observer responses to reverse illusions of different parameters could help perfect models that trick people. Then, we could use the data collected to build models that explain relationships between parameters in our models, and what occurs inside the brains of human observers.

During our investigations, we found that after knowing how an illusion works, or even that a photo is an illusion, we lose our ability to be tricked by this illusion because we implicitly correct the picture 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 the inference map about illumination, which means the spatial continuity of illumination will be decreased, and therefore the illusion will also be decreased in power.

In conclusion, our models explain how human observers are tricked by Simultaneous Contrast illusions and the Adelson illusion. However, the models still have the potential to include related phenomenon and be more explicit. This could be completed with further study. These models could then help us better understand and explain how people infer color from 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

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

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

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.

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

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

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

Goodman, N. D., & Stuhlmüller, A. (2014). The Design and Implementation of Probabilistic Programming Languages (electronic). Retrieved from http://dippl.org.

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