Understanding Human Perception Through Mooney Faces

by

Riya Arora

B.S., Computer Science and Engineering, Massachusetts Institute of Technology (2023)

Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Riya Arora
Department of Electrical Engineering and Computer Science
May 19, 2023

Certified by: Max Siegel
Postdoc in Brain and Cognitive Sciences
Thesis Supervisor

Certified by: Bernhard Egger
Professor of Cognitive Computer Vision
Thesis Supervisor

Certified by: Joshua B. Tenenbaum
Professor of Computational Cognitive Science
Thesis Supervisor

Accepted by: Katrina LaCurts
Chair
Master of Engineering Thesis Committee

Understanding Human Perception Through Mooney Faces

by

Riya Arora

Submitted to the Department of Electrical Engineering and Computer Science
on May 19, 2023, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Human vision is remarkably tolerant to image distortions: even when every pixel in an image has been destructively altered, as in classic Mooney displays, humans can still extract information about identity, pose, and more. Most current deep learning computer vision models perform well with standard face images, but they struggle with stimuli which differ from their training data, like Mooney faces. What makes human perception so comparatively robust? We consider a version of the analysis-by-synthesis proposal for perception, in which visual input is interpreted by inverting a model of image formation, as a potential model for human visual perception. Taking Mooney faces as a case study, we evaluate the model against human performance in a test domain, determining head pose, with the objective of replicating human perception. Previous human psychophysical studies have identified an illusion in which the perceived pose of a Mooney face differs from the pose recovered from an uncorrupted image. The analysis-by-synthesis model does not show a similar effect.

Thesis Supervisor: Max Siegel
Title: Postdoc in Brain and Cognitive Sciences

Thesis Supervisor: Bernhard Egger
Title: Professor of Cognitive Computer Vision

Thesis Supervisor: Joshua B. Tenenbaum
Title: Professor of Computational Cognitive Science

Acknowledgments

I would like to thank my family, especially my parents, for their unconditional love and their constant support for everything I do in life. My parents push me to do so much more than I would otherwise do, and without them I would not be pursuing my MEng.

I would like to thank Joshua Tenenbaum for the opportunity to work in his lab for both my previous UROP and my MEng.

I would like to thank Max Siegel and Bernhard Egger for their support and mentorship throughout my MEng. They were always there for me to teach me how to do research, as well as to support me when anything went completely wrong with my project.

I would like to thank all of my teachers and professors in both grade school and MIT for instilling in me a love of learning and for supporting me in my academic journey.

I would like to thank all the MIT administrators who put together an incredible MEng program.

I would like to thank S^3 who helped me on numerous occasions when things didn't go according to plan and I wasn't feeling well.

I would like to thank all of my mentors along the way, who showed me I could do and be more than what I thought was possible, and who never stopped believing in me.

I would like to thank all my friends and peers at MIT, especially Next House 5E, for incredible experiences, helping me out with psets when I was struggling, and for being some of the best friends in the world.

I wouldn't be where I am today without everyone listed above and many more, so thanks to everyone who was a part of my journey and who brought me to where I am now.

Contents

List of Figures

List of Tables

1. Introduction

Perceiving faces is a complicated tasks that humans are able to do almost instantly. Being able to recognize faces is critical for people in day to day interactions. Furthermore, analyzing faces helps people determine expressions and emotion. Human facial perception is one of the fundamental aspects of our social and cognitive development, and it's a fundamental part of being human. Therefore, understanding how humans perceive faces has significant implications for a range of fields, including psychology, neuroscience, computer vision, and artificial intelligence.

In this thesis, we take human face perception as a case study of object perception and human vision more generally. Using the domain of Mooney faces, or extremely degraded images of human faces, we evaluate whether an analysis-by-synthesis model can explain human percepts which are not well matched by current deep learning models. To do so, we develop a method for generating Mooney versions of standard color images, we measure human performance in perceiving these Mooney faces, and we test whether the model outputs similar percepts as reported by people.

1.1 Mooney Faces

A Mooney image is a distorted, black and white representation of a typical color image. The image can show anything, including animals and objects. The most commonly used Mooney images show human faces. Mooney faces are an important tool in studying human vision: they are extremely degraded versions of facial images, which humans can nevertheless still perceive as human faces. They were first studied in [9]. As seen in Figure 1-1, the three Mooney images on the right contain a fraction

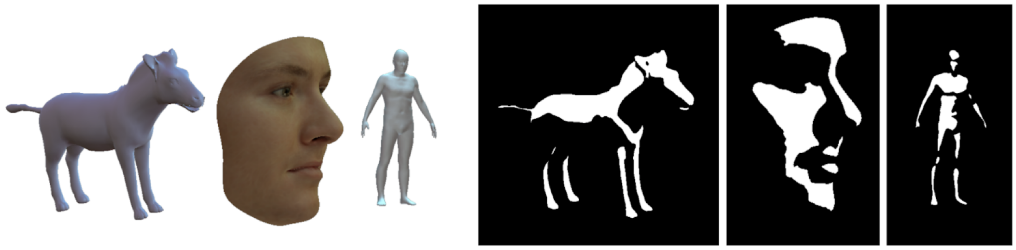

Figure 1-1: The left is an example of three RGB images and the right has the corresponding three Mooney images.

Figure 1-2: The left is an example of an RGB face and the right has the corresponding Mooney face.

of the information of the original RGB images on the left, but we can still see that the 3 original images were a horse, a face, and a person.

Another example of a Mooney face is in Figure 1-2. It is quite surprising that humans can still recognize the Mooney face as a face given that it looks very different from the original face. We hope that by understanding the mechanism by which humans recognize Mooney faces, we will get a deeper understanding of human facial recognition.

In most other work, Mooney faces are generated manually. However, [6] gave a neural network mechanism to generate Mooney faces, discussed in detail in their paper. Here, we used a simpler algorithm to generate Mooney faces [4]: First we blur the image, and then we set the brightest 25% of pixels to be white and the remaining 75% of pixels to be black.

1.2 Generalization in Human Perception

Humans are able to generalize across different stimuli and group images together using these generalizations. To understand and reflect human performance, our model should mirror this generalization ability.

In current literature, human perception is often modeled using deep learning [15, 13]. Deep learning models need to be trained on a particular type stimuli before they can perform well on that type of stimuli. Deep learning models that have not been trained on Mooney faces therefore perform badly at perceiving Mooney faces (for example, in Table 1 of [4], the normalized accuracy of the EIG model from [15] is 0.00). In order for deep learning models to perform well on atypical stimuli, the models might need to be trained on all types of atypical stimuli that could potentially show up, which would take a lot of data.

Furthermore, these models are not able to simulate the human behavior of generalizing to images which have never before been seen, as noted in [4]. Humans can perceive Mooney faces as faces without having any prior exposure to Mooney faces. Deep learning models do not have this ability to perceive Mooney faces without having seen Mooney faces before in their training data. To address this deficiency, in this project we test an alternative type of model, a "top down" inverse rendering model incorporating models of face geometry and lighting, which may enable such perceptual generalization, and which we believe is closer than deep learning to the way humans perceive faces. Our goal is to get a better understanding of human perception of faces by studying human vision in the extreme case of Mooney faces. We attempt to mimic human perception of Mooney faces with a top down inverse rendering Markov Chain Monte Carlo (MCMC).

1.3 Measuring Human Perception

To measure how humans perceive Mooney faces, we created an experiment where participants were given 2 Alternate Forced Choice (2AFC) tasks. In these tasks, participants where given two options in response to different types of questions and had to pick one of the two options. The specifics of the 2AFC experiments we ran are discussed in detail in the sections corresponding to the different experiments.

2. Related Work

Mooney faces were originally developed as reduced representations of faces [9]. In the original work, people were asked to complete the Mooney Test, which involved looking at Mooney faces and determining the age and gender of the original face[9]. Another age detection task was conducted using Mooney faces in [1]. Mooney faces and Mooney images have also been used for other tasks, such as determining if a Mooney face is a ”true“ face or not, or attempting to recognize an object from a Mooney image. Our tasks differ from these works in that we ask participants to match a face to a Mooney face in the first experiment and we ask participants to determine the angle of a face in the second experiment.

Previous work used a similar setup to our first experiment, asking humans to pick which of 8 Mooney faces maps to the original face and comparing that to an inverse rendering model [4]. However, this 8AFC task was difficult for humans and the people doing the experiment may not have been sufficiently incentivised for accuracy. The participants recruited via Amazon Mechanical Turks performed close to randomly, while researchers in the lab, who spent a lot of effort doing this task accurately, had a significantly higher accuracy. For those reasons we believe that the results of that experiment, and the accuracy the authors of [4] got from their participants, may not reflect maximum human capabilities. Also, in that earlier work, the computational method used to evaluate the inverse rendering model was unsatisfactory; rather than having to perform an identical task to what humans performed, for simplicity the model was given the correct pose of the face rather than needing to infer it. We will be building upon this work for our first experiment. We will use a 2AFC task instead of an 8AFC task in hopes that participants will find the task easier and participants outside the lab will get better scores. We will also remove the requirement that the inverse rendering computational model be given the correct pose of the face.

Another previous experiment demonstrated an illusory effect of illumination: when illuminated from different angles, a grayscale face can appear to take on different poses. Participants were presented with two such faces and asked to determine if the second image was rotated to the right or to the left of the original image [12]. We will be building on this work in our second experiment, with a few differences, in order to test whether our computational model can mimic the effect that these faces have on humans. We will also make a few smaller changes, such as using Mooney faces instead of grey scale faces, and running a slightly different 2AFC task.

Computational models to study human face perception have included generative models and inverse rendering (i.e. analysis-by-synthesis) models [8]. However, that work studied only normal RGB images. In our work, we will use the extreme case of Mooney faces to study human facial perception.

Several recent papers have compared computational models of face processing with human perceptual judgments. One approach is to use neural networks to model human perception [13]. Another approach, as in the EIG model [14], uses deep learning to mimic the neural computations that animals go through when processing images. The issue with both of these approaches, for our purpose, is that these models are typically tested on similar data to the dataset they were trained on. This means that the model does well on all the types of images present in their given dataset, but would perform poorly on (for example) Mooney faces, as Mooney faces aren't present in their training dataset [4]. This is unlike human behavior, as humans have also never seen Mooney faces before but perform better than the EIG model on the 8AFC task to pick a Mooney faces that corresponds to an RGB face from 8 Mooney faces [4].

There is also neuroscience work attempting to link the neural responses facial features a person sees [2] to algorithmic processes. However, this work also did not consider different types of abnormal stimuli, such as Mooney faces. There is work which uses Mooney objects – the Mooney transform applied to non-face stimuli - to study object perception, but it did not examine faces [16].

Other work has focused on computational approaches to reconstructing a 3D face from a Mooney face [7, 11]. However, this work does not include human experiments and it does not propose a computational model for facial recognition of Mooney faces. Like this work, however, we also generate faces from a parametric face representation [5].

Most previous work generates Mooney faces manually. One way to generate Mooney faces using a non manual method is to use a deep learning, neural network approach [6]. There is also work on a simpler approach where we take the normal RGB face, blur it, and make the brightest pixels white and the rest of the pixels black [4]. We will be using this simpler approach to generate Mooney faces from RGB faces in this work.

Specifically, we generated Mooney faces using the following approach. First, we blur the image using a Gaussian blur with a filter- size of 5x5 pixels with $\sigma = 35$. We then used a threshold that made the brightest 50% of pixels in the image white and made the rest of the pixels black.

3. Computational Model

We want to design a computational model to solve the 2AFC task in a way that mimics human performance as closely as possible.

Given an image of an RGB face, we want to generate a synthetic face[5] that most closely matches that RGB face. We make use of a recently developed inverse rendering model [10] that was built upon in [4]. The model generates images by modeling the physical processes, involving objects, materials, and light, which produce images. It then compares these images with input data to try to interpret them. For example, when given an input Mooney face at a certain pose, the model hypothesizes that the image represents various face shapes, lighting directions, and face poses, and generates images corresponding to these choices. It then determines which generated image is closest to the input Mooney face, and estimates that the Mooney face has the shape, lighting, and pose parameters of that generated image. The model uses an algorithm called Markov Chain Monte Carlo (MCMC) to compute its estimates.

We started with the inverse rendering model in [4], which is given the pose of the RGB face it starts with and then takes 10,000 MCMC samples to infer a synthetic face, including illumination, which matches as closely as possible the original face. For the Mooney faces, we used the *IlluminationOnly* method described in [4]. This method treats Mooney faces like RGB faces, and when the MCMC model is run, the result has illumination which results in a very black and white image. An example of the model's result from fitting a Mooney face is in Figure 3-1.

We made a few changes to this model for our work. First, instead of giving the correct pose to the model, we wanted the model to infer the it. Humans are not given the pose; they estimate it from the images. We wanted our model to do the same. To

Figure 3-1: This is an example of the model being fit to a Mooney face. The left face is the original Mooney face and the right face is the fit RGB face, which then yields an image which is very black and white.

do this, we gave the model n possible initial yaw angles and did not allow it to vary the yaw angle of the face that we had given it. For each of the n initial yaw angles, we ran the model with 2,000 MCMC samples to get the best guess of what the generated face would look like with that yaw angle. This gave us multiple possible generated faces, one for each initial yaw angle. We then picked which of the generated faces has the smallest L2 norm distance to the target face as the generated face which is the best fit to the target face.

The following equation is a mathematical representation of how we pick which of the n generated faces is the best fit to the target Mooney face. Here, $nRows$ and $nCols$ are the number of rows and columns respectively of pixels in each image. $T(i, j)$ is the pixel at th ith row and jth column in the target Mooney image. G_x corresponds to the xth generated face out of the n generated faces. ".r", ".g", and ".b" corresponds to the r, g, and b value of the given pixel respectively.

$$\operatorname*{argmin}_{x \in [1,n]} \sqrt{\frac{\sum_{i=1}^{nRows} \sum_{j=1}^{nCols} (T(i,j).r - G_x(i,j).r)^2 + (T(i,j).g - G_x(i,j).g)^2 + (T(i,j).b - G_x(i,j).b)^2}{3 * nRows * nCols}}$$

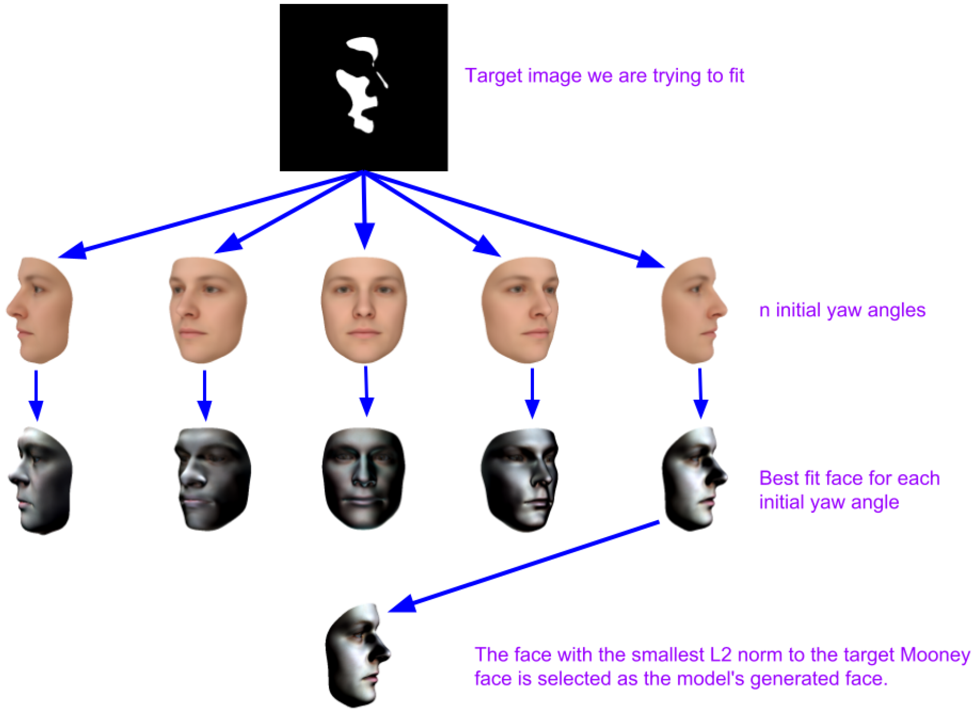

Figure 3-2: Diagram of all the steps the model algorithm goes through to find the best fit generate face for a given Mooney face.

This method provides a very close pose, if not the perfectly correct pose, for a randomly generated Mooney face. Figure 3-2 summarizes the process that the model goes through to determine the best fit generated face for a given target face.

3.1 Initial yaw angles for the original experiment

In the initial experiment, the Mooney faces given could be facing either left or right, so the initial yaw angles had to be in both direction. There were $n = 9$ initial yaw angles, spaced evenly between -90 degrees and 90 degrees. Figure 3-3 shows a Mooney faces and all $n = 9$ generated faces from the initial yaw angles. The big face on the bottom left is the best fit face as picked using the method described above.

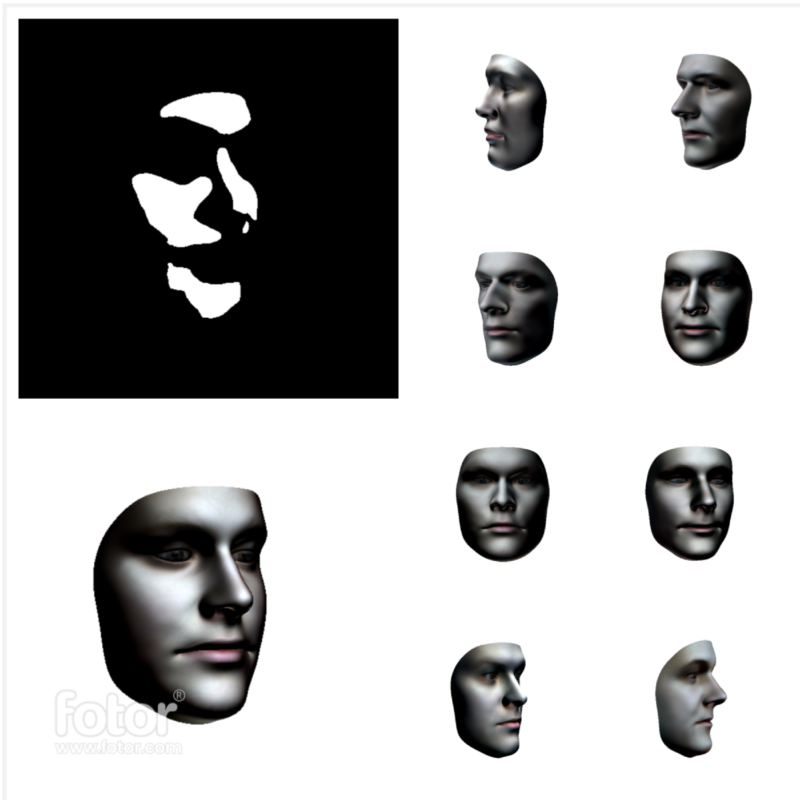

Figure 3-3: The top left image is the Mooney face we are trying to fit. The model tries $n = 9$ different starting yaw angles. The generated faces for each of the initial yaw angles are the $n = 9$ other faces shown in the image. The bottom left face is the best fit face as determined by the model. The 8 small faces to the right are the other 8 generated images from different starting yaw angles.

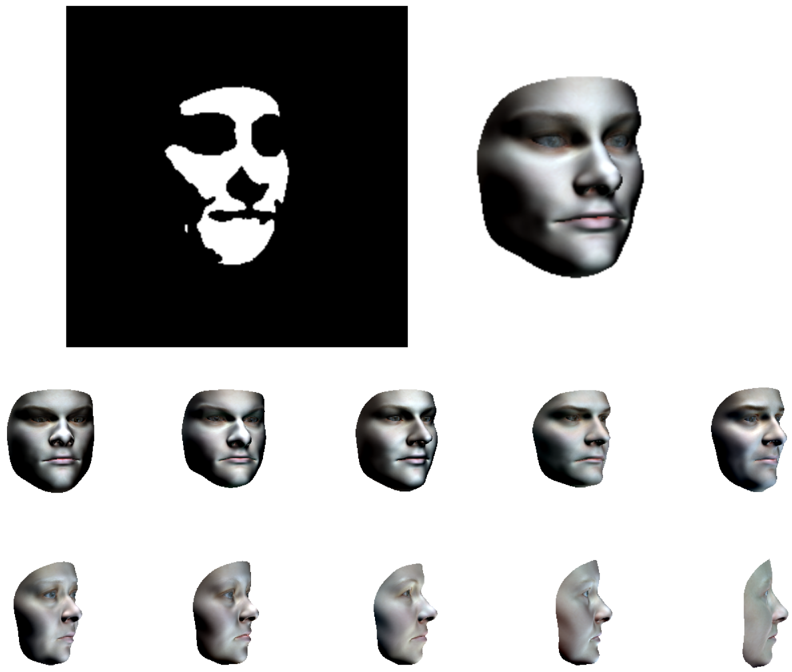

Figure 3-4: The top left image is the Mooney face we are trying to fit. The model tries $n = 11$ different starting yaw angles. The generated faces for each of the initial yaw angles are the $n = 11$ other faces shown in the image. The top right face is the best fit face as determined by the model. The 10 small faces on the bottom to the right are the other 10 generated images from different starting yaw angles.

3.2 Initial yaw angles for the new experiment

In the new experiment, the Mooney faces given could only be facing towards the right, so the initial yaw angles where only pointed towards the right. In this experiment, the accuracy of the guessed yaw angle mattered significantly more than in the original experiment, so we had more initial yaw angles every 7.5 degrees instead of every 20 degrees in the initial experiment. There were $n = 11$ initial yaw angles, spaced evenly between 15 degrees and 90 degrees. Figure 3-4 shows a Mooney faces and all $n = 11$ generated faces from the initial yaw angles. The big face on the top right is the best fit face as picked using the method described above.

3.3 Answering the 2AFC task

We then need to determine which of the two options of the 2AFC task is the better option. The methods for doing this are different for the 2 different experiments we tried running, and are described in depth with each of the experiments in sections 4.2 and 5.2.

4. Original Experiment

For the initial MEng proposal, we suggested an experiment which did not yield informative results. We tried many different ways of modifying the experiment in hopes that we would get better results but none of those ways worked. In this chapter, I will discuss what the original experiment was, the various modifications which, we tried to get it to work, and why we think it did not end up working.

4.1 Overview of the experiment

We started by taking the 8AFC task reported previously and improving on it. Recall from Chapter 2 that the task involved giving participants one RGB face and 8 Mooney faces where one of the Mooney faces was a Mooney face of the original RGB face[4]. The participants then had to pick which of the Mooney faces matched the original RGB face. An example of the 8AFC task is shown in Figure 4-1.

The reason the authors of [4] believed that task didn't do to well was because it was too difficult. This caused unmotivated participants to perform close to chance while motivated participants from the lab did significantly better. This made the results unreliable.

To improve upon this, we created a 2AFC task similar to their 8AFC task in hopes that the task will be significantly easier if participants only have to pick from 2 Mooney faces instead of from 8 Mooney faces. The idea was that because the task is significantly easier, participants will be more motivated and get a higher accuracy. An example of the 2AFC task is shown in Figure 4-2.

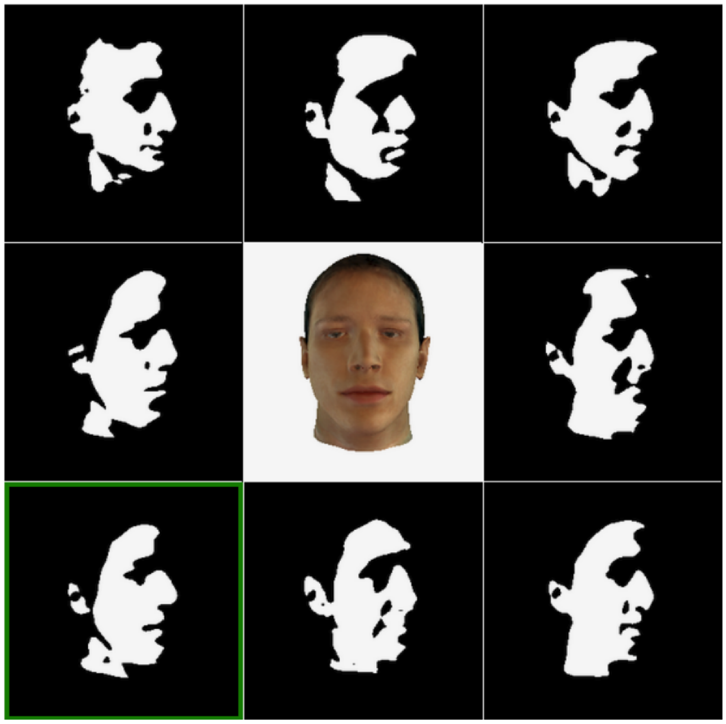

Figure 4-1: This is the 8AFC task done by previous researchers. The image in the green outline is the correct match to the RGB image.

Figure 4-2: Example of the first 2AFC experiment we tried. The image in the green outline is the correct match to the RGB image.

4.2 Computational Model

Recall from Chapter 3 that for each Mooney face, we infer the parameters of the face that which most closely matches the face shown in an input image. Now that we have a model for matching a face based one *one* image, we need to have the model answer the 2AFC task.

To do this, we run the model to generate a face for all three images: the RGB face that we are trying to guess and the two Mooney faces which are our options. We then take the first 10 shape parameters of the model we generated of the face we are trying to guess and take the cosine angle between that and the first 10 parameters of the shapes each of the two option faces. We only take the first 10 parameters of the shape because those matter the most as to the shape of the face. More details on why we only take the first 10 parameters of the shape of the face in section 4.3.2. The option which has a bigger cosine angle of its shape to the shape of the face we are trying to guess is what we select as the option that the model picks.

The bigger the difference between the two cosine angles, the easier the task is for the model, and the more sure the model is of its answer.

4.3 The experiments we ran

The participants were shown instructions detailing how to respond to the trials followed by 100 trials similar to the one shown in Figure 4-2. In addition to the detailed instructions at the beginning, each trial had a summary of how participants should their response to the trial on the bottom. Each trial had an RGB face on top and two Mooney faces on the bottom. One of the Mooney faces corresponds to the RGB face. This means that the RGB face was taken, a random illumination was placed on it, it was turned to a random angle, and then it was made to be a Mooney face. The correct Mooney face was placed at random on the left or on the right of the two bottom Mooney faces. The participants would click the left arrow key to signify that the Mooney face on the left matches the RGB face or they would click the right arrow key to signify the Mooney face on the right matches the RGB face. The faces, which means the color and shape of the face, that were generated at random were generated from the Basel Face Model [5]. The random illuminations were generated at random from the Basel Illumination Prior [3].

The participants were forced to wait 750 ms before entering their response. This is enough time to ensure that they don't click through the trials without trying, but short enough to ensure that it does not effect or hinder a participant putting effort into getting the correct answer.

4.3.1 Our first 2AFC experiment

In the original 8AFC experiment, the center RGB face is facing forward and has no illumination on it. Each of the 8 Mooney faces are facing at a 60 degree angle to the right and all of the Mooney faces have the exact same illumination shined on them before the Mooney images were created. In all the tasks of the experiment, the illumination shined on the Mooney faces before the Mooney images were created is identical. An example of this is shown in Figure 4-1.

In our 2AFC experiment, we made several changes from this. While the RGB face remained facing the front with no illumination, but each of the Mooney faces had a random illumination and was facing a random angle. An example of this is shown in Figure 4-2.

We then ran this experiment on 5 participants and they scored an average of 60.2% on this task. Guessing will give a score of 50%, so the participants performed only slightly better than chance. These 5 participants were motivated and when we run the experiment on a massive amount of unmotivated people we expect the performance to decrease. This means that if we were to run this experiment on more people to get data, the results will be even closer to chance and harder to use reliably. We then tried several things to make the experiment easier and improve human performance so that we can get better data that is much further from chance.

4.3.2 Different shape of the face

One option we tried to make the tasks easier is to pick two underlying faces with vastly different shapes from each other. To determine faces that were different, we took the shape parameters from [3] for each of the two faces and took the cosine angle between the two. One would expect that when this cosine angle is close to 1, the faces would have a very similar shape, and when the cosine angle is close to -1, the faces will have a very different shape. However, upon visual inspection of the pairs of faces, it wasn't the case that a higher the cosine angle meant a more similar shape. The cosine angle seemed unrelated to how similar or different the shapes of the faces were.

The reason for this was that the first few parameters of the shape impact the shape of the face a lot, and the later parameters effect the shape less and less. To fix this, we took the first 10 parameters of the shape for each face and took the cosine angle between these. Now the pairs of faces with a cosine angle closer to 1 looked like they had similar shapes and the pairs of faces with a cosine angle closer to -1 looked like they had different shapes. Figure 4-3 shows an example of 2 faces with a cosine angle close to -1. These faces have different shapes. Figure 4-4 shows an example of 2 faces with a cosine angle close to 1. These faces have similar shapes.

We then took the pairs of faces with different shapes to create the tasks and thought that these would be easier than a random pair of faces. We picked these faces by randomly generating 1,000 faces and picking pairs of faces with very negative cosine angles. People performed slightly better on these tasks, averaging 65% on 5 motivated participants in the lab, but still very close to chance.

Another thing we tried was mimicking the illumination and pose conditions of the original experiment (we describe in detail how we do this in section 4.3.6) and creating faces where the shape was different. We did this by randomly generating one face, and then multiplying all the shape parameters where the parameter had a small absolute value by 2 to create one face. The other face was created by taking the first face and multiplying all the shape parameters by -1. Participants in the lab performed very well on these tasks, averaging 80%. However, the participants were

Figure 4-3: An example of 2 faces with a different shape

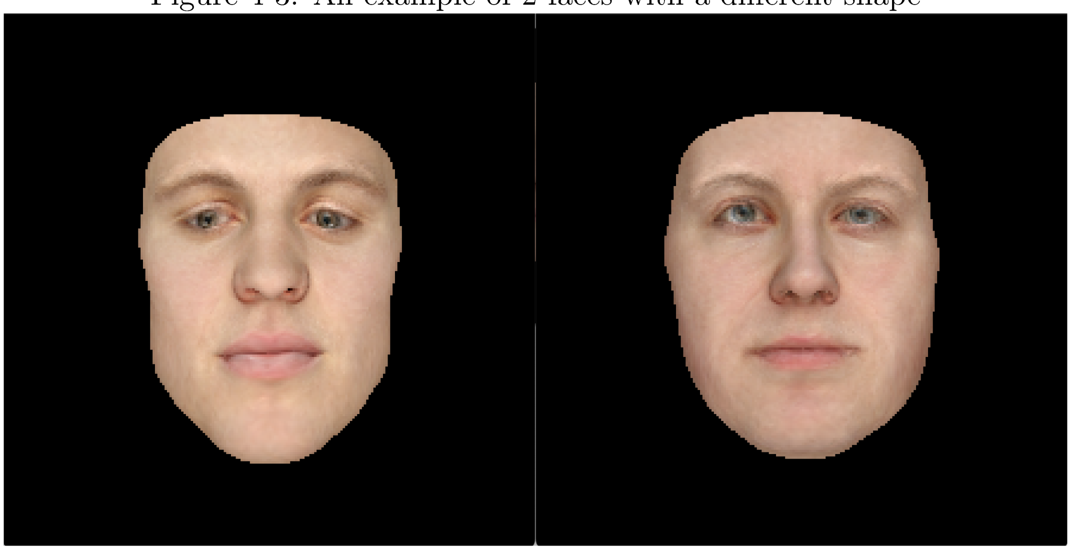

Figure 4-4: An example of 2 faces with a similar shape

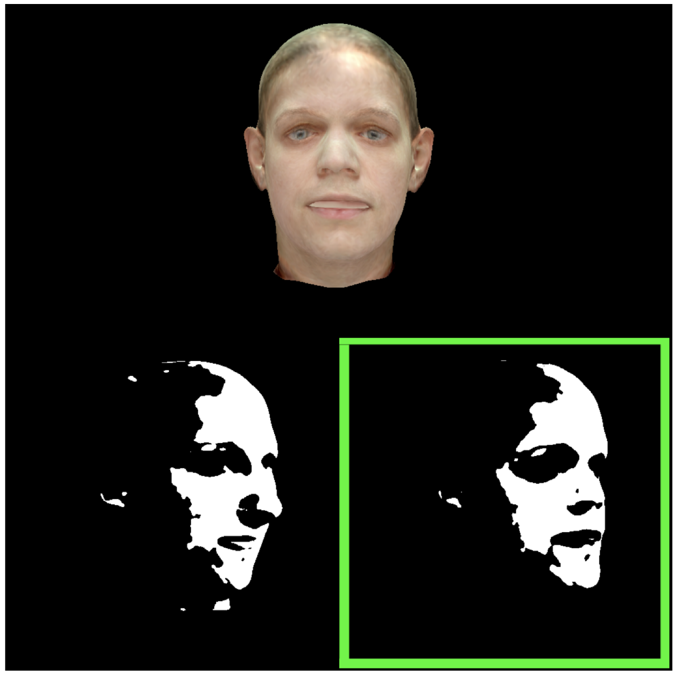

Figure 4-5: This is an example of a 2AFC task where we specifically generated the faces to have different shapes. The image in the green outline is the correct match to the RGB image. The answer can be easily determined by looking at the mouth here. This example also mimics more of the original experiment conditions where both the Mooney faces have the same illumination and are turned at a 60 degree angle to the right.

solving the tasks by looking at certain features of the face such as the mouth, the nose, or the forehead size instead of perceiving the face as a whole. An example of a task like this where it can be answered simply by looking at the mouth is in Figure 4-5. This was less of a facial recognition task and more of a feature matching task, which is not what we were looking for. For that reason, this task wouldn't work either.

4.3.3 Different color of the face

Similar to section 4.3.2 where we took the first 10 parameters of the shape from 2 different faces and took the cosine angle between them, we tried taking the first 10 parameters of the color for 2 faces and taking the cosine angle between them.

What we found was that the illumination that we put over the face also contributes

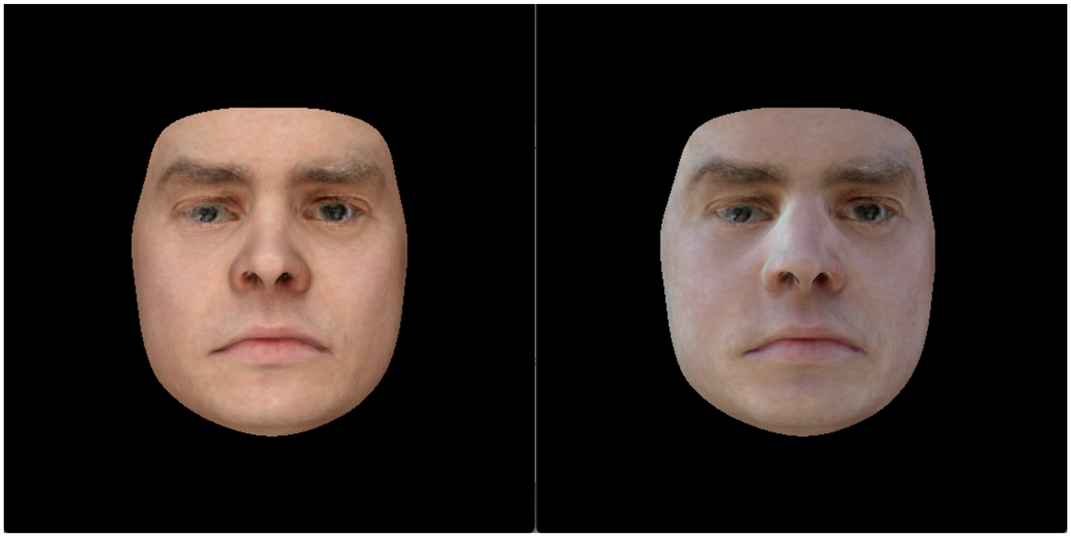

Figure 4-6: The left face is an RGB face with no illumination and the right face is illumination added onto the left RGB face. The illumination changes the color of the face.

to the color, so just looking at the color did not give better results than random faces. Figure 4-6 shows an example of taking a face with no illumination and putting illumination on it to visualize the color change. The left face is an RGB face with no illumination and the right face is illumination added onto the left RGB face. As you can see, the illumination drastically changes the color.

4.3.4 Both Mooneys have a yaw angle more than 70 degrees

Another thing we tried was seeing if looking at the faces from the side. The idea here is that we can see more features from the side of the face than we can from the front. For this experiment, we took faces where the yaw angle of the Mooney face was turned a random number between 70 and 90 degrees either to the left or to the right and the RGB face was still facing forward with a yaw at 0 degrees. An example of this task is show in Figure 4-7. This resulted in motivated participants in the lab scoring 65%.

We also tried making the RGB face face a random angle between 70 and 90 degrees either to the left or to the right. In this case, we also illuminated the RGB face with

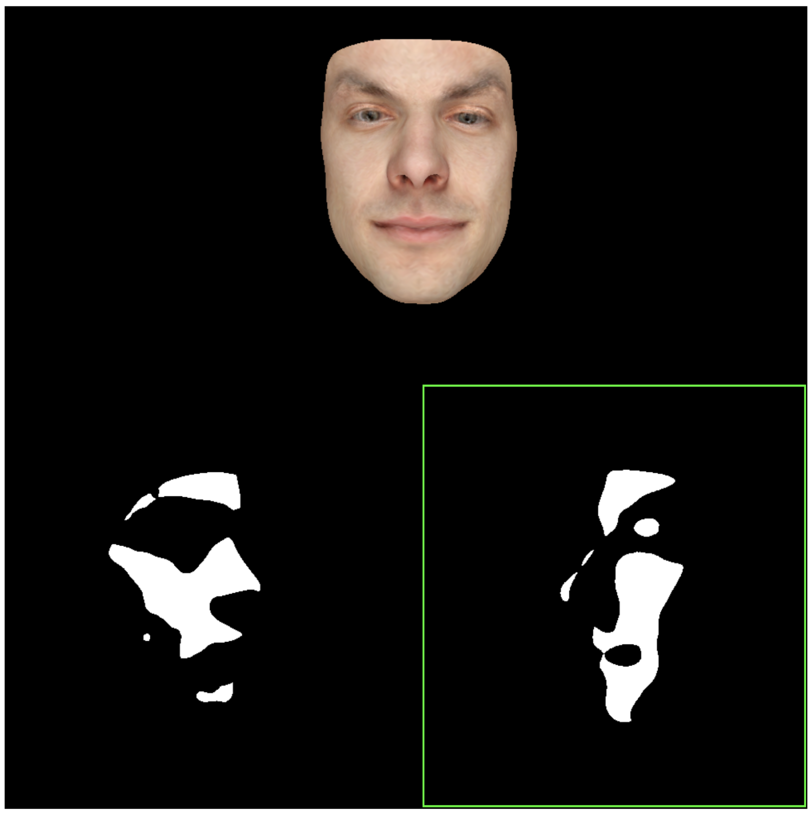

Figure 4-7: This is an example of the task shown where the two Mooney faces are facing at a random direction and with a random angle between 70 and 90 degrees. The correct answer is boxed in green.

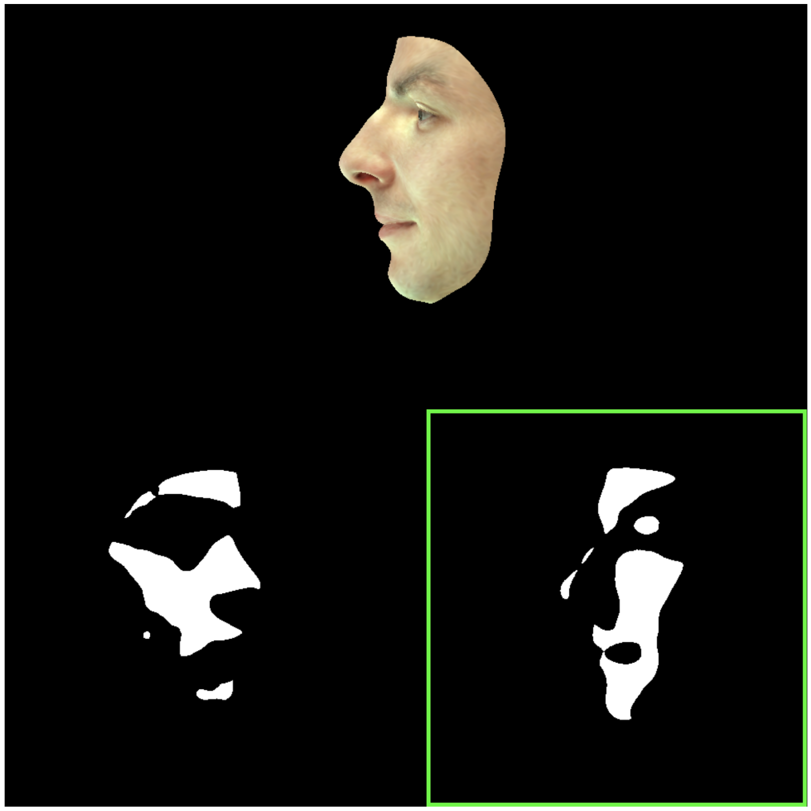

Figure 4-8: This is an example of the task shown where the two Mooney faces and the RGB face are all facing at a random direction and with a random angle between 70 and 90 degrees. The RGB face is also illuminated. The correct answer is boxed in green.

a different random illumination than each the Mooney faces. An example of this task is show in Figure 4-8. This resulted in motivated participants in the lab still scored 65%.

We then combined this approach with the approach where we picked different shapes from section 4.3.2. Here the faces were picked by randomly generating faces and picking ones with different shapes as defined in section 4.3.2. We then took these faces and turned them either left or right at random at an angle selected at random between 70 and 90 degrees. With this approach, motivated participants in the lab were scoring close to 70%, but still too close to chance for us to trust results run on unmotivated participants.

4.3.5 Picking trials that our MCMC computational model finds easy

Recall from section 4.2 that there are certain tasks the computational model finds easier than other tasks. Since the goal of the computational model is to mimic human performance, it follows that tasks the computational model finds easy will also be easy for humans.

We randomly generated faces for this task and asked the computational model to provide us with the difficulty of the task. Recall from chapter 3 that easier tasks have a greater absolute value of the difference in the cosine angles. We then selected tasks which the model found really easy, regardless of if the model got the right answer or not. On these tasks, 10 motivated participants in the lab had an average performance of 63%.

We then slightly changed the experiment to take tasks that the model found easy and the model got correct. On these tasks, 10 motivated participants in the lab had an average performance of 66%. In both of these cases, humans didn't perform much better than chance.

4.3.6 Both Mooneys have the same illumination and a 60 degree angle

In the 8AFC experiment done previously, 6 participants in the lab had an average score of 44.2% (we got this number by asking the authors of [4]), significantly above chance of 12.5%. For this reason, we also tried to mimic the details of the 8AFC task exactly but with 2 Mooney faces instead of 8. For this, we gave each of the Mooney faces a yaw angle of 60 degrees to the right and both of the 2 Mooney faces came from the same illumination, but each tasks had a different illumination. An example of this task is shown in Figure 4-5.

Note that this is different from the original 8AFC experiment where all tasks had the same illumination. However, we believe that the integrity of the experiment goes up if each task has a different randomly generated illumination since that is more realistic than picking one illumination for all the tasks.

The results of the motivated human participants in the lab for this experiment was also close to 65%.

4.4 Why none of these experiments had a high human accuracy

In the 8AFC experiment done previously, 6 participants in the lab had an average score of 44.2%. This lead us to believe that human participants in the lab for the 2AFC experiment would also perform significantly above chance, much higher than the results we were getting of 65% - 70%. However, this was not the case. The most likely theory as to why is that people either recognize a face or they don't recognize it. If a person recognizes the face, it doesn't matter if there is one or seven distractors. They will get it right. Similarly, if a person doesn't recognize the face, they will make a guess. The odds that their guess is right in the 2AFC task is 50% and the odds that the guess is right in the 8AFC task is 12.5% so the overall score is slightly higher in the 2AFC task.

Humans scoring 65% on the 2AFC task implies that they are able to solve 30% of the tasks, and for the remaining 70% they guessed and got half correct. This leads to a total accuracy of $30\% + (70\%)/2 = 65\%$. If humans solve 30% of the tasks in the 8AFC task, they will score $30\% + (70\%)/8 = 38.75\%$.

Humans scoring 70% on the 2AFC task implies that they are able to solve 40% of the tasks, and for the remaining 60% they guessed and got half correct. This leads to a total accuracy of $40\% + (60\%)/2 = 70\%$. If humans solve 40% of the tasks in the 8AFC task, they will score $40\% + (60\%)/8 = 47.5\%$.

Since the average score of the 8AFC task is 44.2%, which is in between 38.75% and 47.5%, it supports the theory that people can either answer a task and recognize the face or not. If they can answer it, the number of distractors doesn't matter.

5. New Experiment

Since we were not able to get great results from the original experiment, we shifted directions to a new experiment where we studied the results from [12] on how humans perceive a face to be at a different yaw angle based on the location of the illumination on the face. While [12] used grey scale faces to analyze this effect, we used Mooney faces to study this effect.

5.1 Experimental Design

For this experiment, we created 3 groups of stimuli. The first group has faces that are facing 30 degrees to the right and are illuminated from the front. An example of a face from this group is shown in Figure 5-1. The left face is an RGB face and the right face is the corresponding Mooney face.

The second group of stimuli has faces that are facing 45 degrees to the right and are illuminated from the front. An example of a face from this group is shown in Figure 5-2. The left face is an RGB face and the right face is the corresponding Mooney face.

The third group of stimuli has faces that are facing 30 degrees to the right and are illuminated from the far left. Both the faces from group 1 and group 3 are facing 30 degrees to the right, the only difference is that the faces from group 3 are illuminated from the far left and the faces from group 1 are illuminated from the front. Figure 5-3 shows a face from group 1 and a face from group 3 side by side.

For the 2AFC task, participants were shown one one of the Mooney faces from either group 1, 2 or 3. The were given 2 RGB faces, one facing 30 degrees to the right

Figure 5-1: This is an example of a face from group 1. The face is turned 30 degrees to the right and the illumination is coming from the front. The left face is an RGB face and the right face is the corresponding Mooney face.

Figure 5-2: This is an example of a face from group 2. The face is turned 45 degrees to the right and the illumination is coming from the front. The left face is an RGB face and the right face is the corresponding Mooney face.

Figure 5-3: There is a face from group 1 on the left and a face from group 3 on the right. Both faces are turned 30 degrees, but the face from group 3 is illuminated from the far left while the face from group 1 is illuminated from the front.

Figure 5-4: This is an example of a face from group 3. The face is turned 30 degrees to the right and the illumination is coming from the far left. The left face is an RGB face and the right face is the corresponding Mooney face.

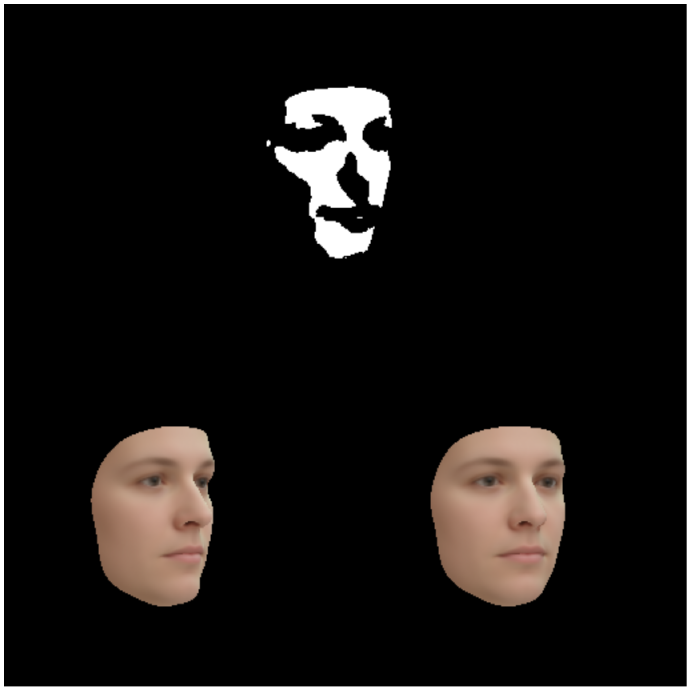

Figure 5-5: This is an example of a the 2AFC task shown to participants.

and one facing 45 degrees to the right, and asked which of the two RGB faces is the Mooney face pointing in the same direction as. The Mooney face was shown on top and the 2 options were shown on the bottom. The order of the two RGB faces on the bottom was random. Figure 5-5 shows an example of the 2AFC task shown to participants. Participants were asked to click the left arrow key if the believed the answer was the left RGB face or the right arrow key if they believed the answer was the right RGB face. The instructions on how to respond to tasks were shown before participants started the tasks and were also shown on every task. Participants were shown 99 such tasks, 33 tasks from each group.

The difference in human perception between the faces in groups 1 and 3 is the effect we are studying. Figure 5-3 shows the effect we are analyzing with RGB images and Figure 5-6 shows the same effect with Mooney images. Group 2 images are included in this experiment because we want to be able to compare the results between participants answers on groups 1, 2 and 3.

Figure 5-6: These are Mooney faces from groups 1 and 3 side by side. Both are turned 30 degrees to the right, but the image from group 3 on the right looks like it is turned more towards the right than the image from group 1 on the left.

5.2 Computational Model

Recall from Chapter 3 and section 3.2 that start with 11 initial yaw angles spaced evenly from 15 degrees to 90 degrees to the right. The MCMC model then finds the best fit face for each of the initial starting yaw angles, and picks the best fit result.

For this experiment, we want to know what the MCMC model guesses as the best initial starting angle. To do this, we take the best fit face from the 11 generated faces and pick the yaw angle of that best fit face. This gives us the yaw angle that our model believes is the angle of the Mooney face.

If we want the MCMC model to answer the 2AFC task, we check if the angle the model determined is closer to 30 degrees or 45 degrees. The one change we make when answering the 2AFC task is that we remove 37.5 degrees as a possible starting yaw angle and replace it with 2 initial starting yaw angles of 35 degrees and 40 degrees. This is because 37.5 degrees is evenly spaced between 30 degrees and 45 degrees, so the model answering 37.5 degrees doesn't give us any information about the model's answer to the task. However, we still want to have possible angles between 30 degrees and 45 degrees, so we add 35 degrees and 40 degrees as possible initial yaw angles.

5.3 Analysis of Results

The experiment was run on 11 motivated participants. Each participant was shown the same set of 99 images. The 99 images contained 33 images from each of the 3 groups. The MCMC computational model also analyzed the same 99 images. However, instead of being asked to solve the 2AFC task, the model was asked to give its best guess of the angle the given Mooney face was turned to.

5.3.1 Human Results

Figure 5-7 summarizes the results of humans participants. For each of the 33 tasks in group, we looked at what percent of the 11 participants answered 45 degrees as the facial orientation and the results are shown in the bar graphs in Figure 5-7. It shows that the participants clearly perceive the stimuli in group 2 to have a higher angle than the stimuli in group 1. It also shows that participants perceive the stimuli in group 3 to have a higher angle than the stimuli in group 1. This result is expected as it is very similar to the results from [12] on grey scale faces. However, humans perceive the stimuli from group 3 to have a smaller angle than the stimuli from group 2.

5.3.2 Computational Model Results

The computational model was asked to give its best guess of the angle of each of the 99 stimuli which had 33 stimuli from each of the 3 groups. Figure 5-8 summarizes the results. There is one bar graph for each of the groups of stimuli. The x-axis was the angle the computational model guessed using the method described in section 5.2 and the y-axis is how many of the 33 tasks resulted in the corresponding angle. Recall that the model is allowed to answer angles in the range from 15 degrees to 90 degrees in increments of 7.5 degrees.

As shown in Figure 5-8, the model generally perceives faces in group 2 to have a higher angle than faces in group 1. This is further supported by Tables 5.1 and 5.2

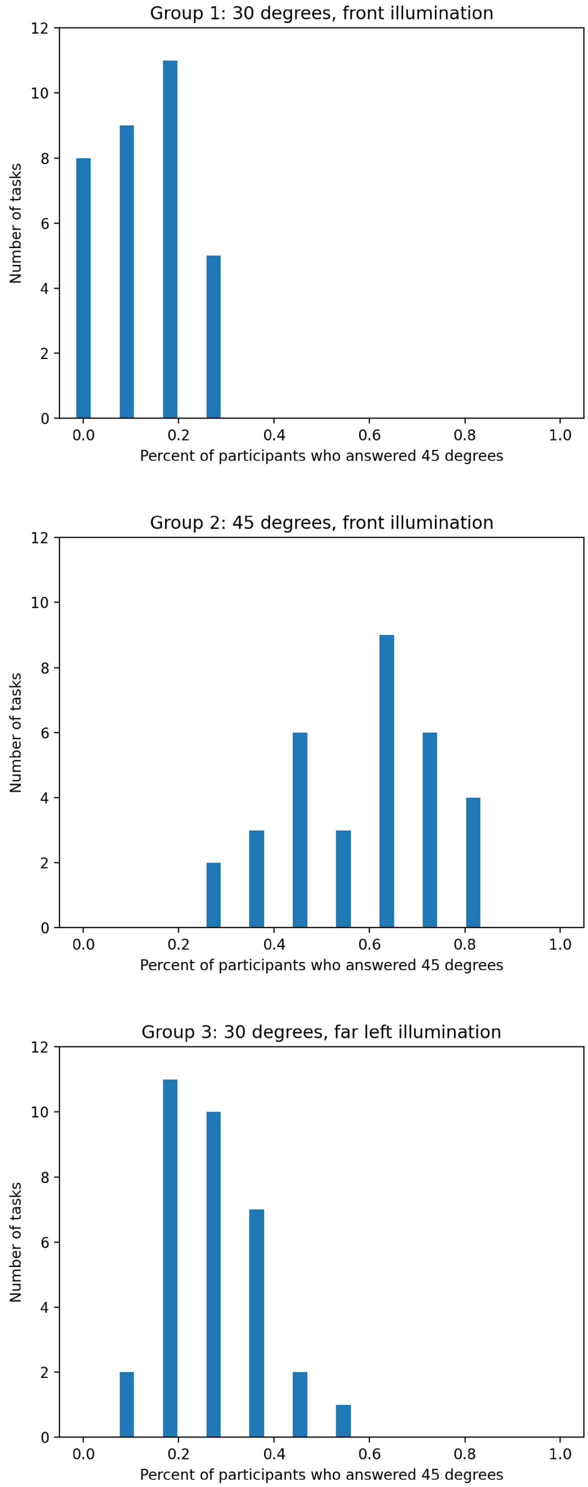

Figure 5-7: The results of the participants on the stimuli in each of the 3 groups.

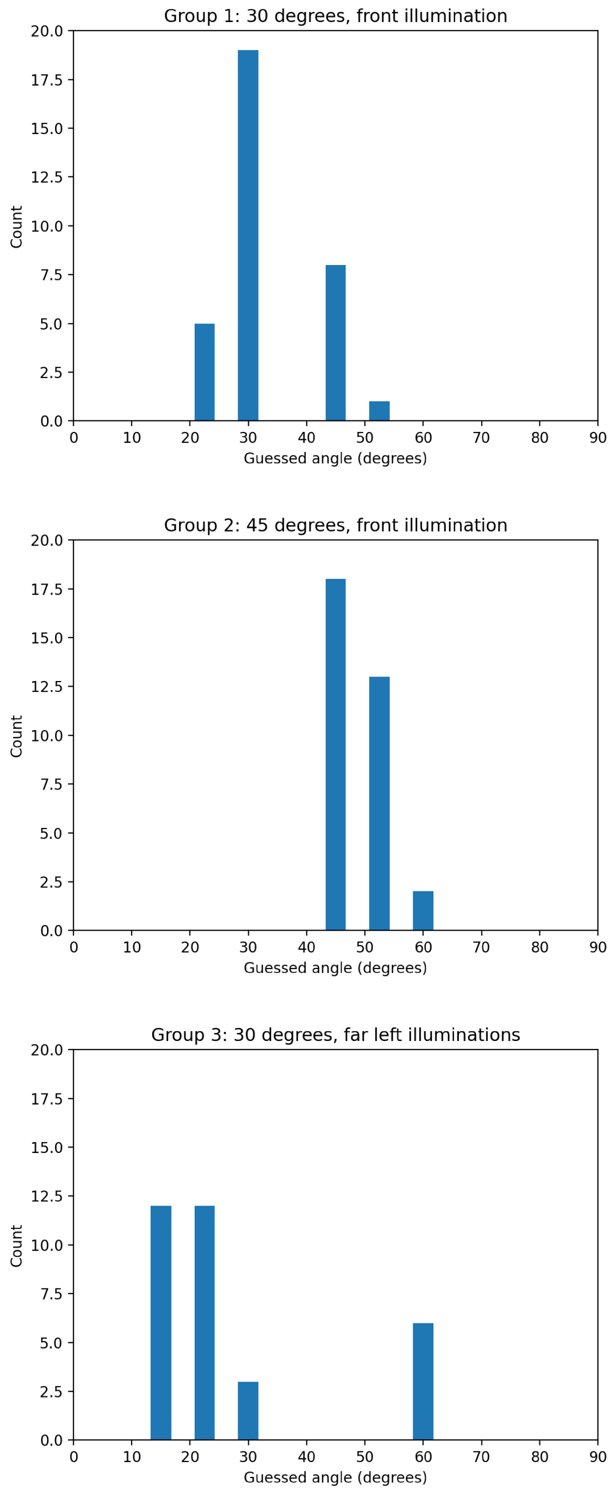

Figure 5-8: The results of the computational model on the stimuli in each of the 3 groups.

Mean Angle (Degrees)		
Group 1	Group 2	Group 3
33.18	48.86	27.27

Table 5.1: The mean angle guessed by the model for each group of stimuli.

Median Angle (Degrees)		
Group 1	Group 2	Group 3
30	45	22.25

Table 5.2: The median angle guessed by the model for each group of stimuli.

where it is shown that the model has a higher median guess and mean guess for the angle of group 2 stimuli than it does for group 1 stimuli. This makes sense as the faces in group 2 have a bigger angle than the faces in group 1 and both groups have the same front illumination.

The result we were looking to build upon from [12] shows that humans perceive a face to be turned further to the right when illumination is shown on the far right of the face. We also had similar results for human data. However, the model seems to have the opposite result. It perceives the faces in group 3 to be turned further to the left than the faces in group 1. This is shown in Tables 5.1 and 5.2 where it is shown that the median and mean guess for the model for group 3 stimuli is smaller than the median and mean guess for the model for group 1 stimuli. This can also be visualized in Figure 5-8.

One possible reason for this reverse effect is that most faces with the face pointing further to the right will have a thinner face visible than the ones presented in group 3, so the model sees a huge number of pixels which are white in the Mooney image but dark in a generated face with a high angle. People may not be looking as closely at how wide the face presented is, but this matters a lot to the way the model decides which generated face to pick. Figure 5-9 shows and example of a Mooney face from group 3 and an example of a Mooney face turned very far towards the left. The Mooney face from group 3 is wider than the one turned to a far angle.

Figure 5-9: The left is an example of a Mooney face from group 3 and the right is an example of a face turned very far to the left.

5.3.3 Comparison of Results

Next, we wanted to analyze each task to see if tasks that humans found to be 45 degrees were the same ones that the model found to have a high angle. Figure 5-10 shows a scatter plot for each group of stimuli. Each point is one of the 33 tasks. Since tasks can have the same coordinates, the size of the point corresponds to how many tasks are at that point. The results are shown in Figure 5-10 is that there is no correlation between tasks that participants thought were 45 degrees and tasks that the model found to have a high angle.

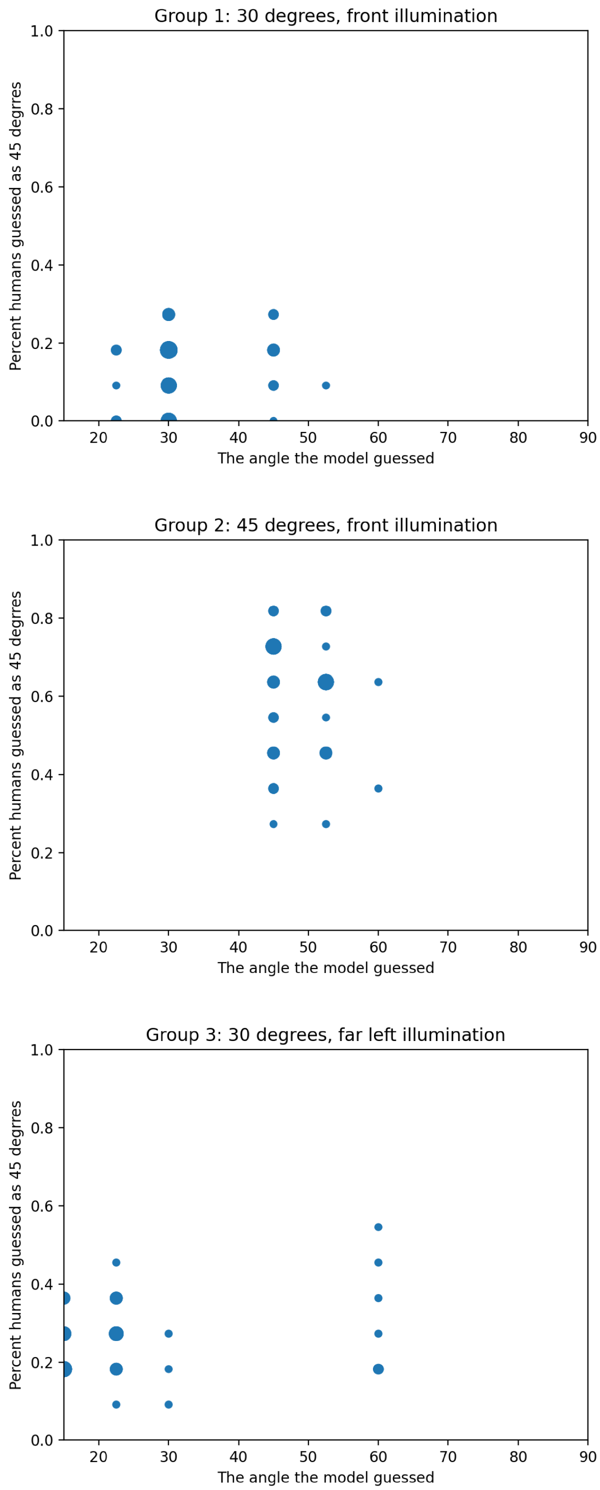

Figure 5-10: A comparison of participants' performance with the computational model's performance in each of the 3 groups.

6. Conclusion

In this project, we studied Mooney faces towards better understanding human face perception, and human object perception more generally. Most facial perception models use deep learning, but deep learning fails to perform well on stimuli it has not been trained on. This includes atypical stimuli such as Mooney faces. We evaluated an alternative analysis-by-synthesis approach as a model for human perception.

In our first experiment, we compared human percepts against model judgments in the domain of human facial perception. However, human performance on the 2AFC task of human facial perception was not high enough to allow a meaningful comparison.

In our second experiment, we looked at a visual illusion in which lighting direction influences human perception of pose. We measured human judgments and compared them with pose inferences from the model While our human data replicated the optical illusion using Mooney faces, our computational model did not show this same effect.

The natural next steps are to improve the computational model or try a different type of computational model to get a model that performs similar to humans on the second 2AFC task of modeling human perception of facial orientation.

Bibliography

[1] Claus-Christian Carbon, Martina Grüter, and Thomas Grüter. Age-dependent face detection and face categorization performance. *PLoS One*, 8(10):e79164, 2013.

[2] Le Chang, Bernhard Egger, Thomas Vetter, and Doris Y Tsao. Explaining face representation in the primate brain using different computational models. *Current Biology*, 31(13):2785–2795, 2021.

[3] Bernhard Egger, Sandro Schönborn, Andreas Schneider, Adam Kortylewski, Andreas Morel-Forster, Clemens Blumer, and Thomas Vetter. Occlusion-aware 3d morphable models and an illumination prior for face image analysis. *International Journal of Computer Vision*, 126:1269–1287, 2018.

[4] Bernhard Egger, Max H Siegel, Riya Arora, Amir Arsalan Soltani, Ilker Yildirim, and Josh Tenenbaum. Inverse rendering best explains face perception under extreme illuminations. *CogSci*, 2020.

[5] Thomas Gerig, Andreas Morel-Forster, Clemens Blumer, Bernhard Egger, Marcel Luthi, Sandro Schoenborn, and Thomas Vetter. Morphable face models - an open framework. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 75–82, 2018. Acknowledging Prof. Dr. T. Vetter, Department of Computer Science, and the UNIVERSITY OF BASEL as the source of the DATA stemming from the Basel Face Model.

[6] Tsung-Wei Ke, X Yu Stella, and David Whitney. Mooney face classification and prediction by learning across tone. In *2017 IEEE international conference on image processing (ICIP)*, pages 2025–2029. IEEE, 2017.

[7] Ira Kemelmacher-Shlizerman, Ronen Basri, and Boaz Nadler. 3d shape reconstruction of mooney faces. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[8] Aleix M Martinez. Computational models of face perception. *Current directions in psychological science*, 26(3):263–269, 2017.

[9] Craig M Mooney. Age in the development of closure ability in children. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 11(4):219, 1957.

[10] Sandro Schönborn, Bernhard Egger, Andreas Morel-Forster, and Thomas Vetter. Markov chain monte carlo for automated face image analysis. *International Journal of Computer Vision*, 123:160–183, 2017.

[11] Amnon Shashua. On photometric issues in 3d visual recognition from a single 2d image. *International Journal of Computer Vision*, 21(1-2):99–122, 1997.

[12] Nikolaus Troje and Ulrike Siebeck. Illumination-induced apparent shift in orientation of human heads. *Perception*, 27:671–80, 02 1998.

[13] Daniel LK Yamins, Ha Hong, Charles F Cadieu, Ethan A Solomon, Darren Seibert, and James J DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, 111(23):8619–8624, 2014.

[14] Ilker Yildirim, Mario Belledonne, Winrich Freiwald, and Josh Tenenbaum. Efficient inverse graphics in biological face processing. *Science advances*, 6(10):eaax5979, 2020.

[15] Ilker Yildirim et al. Efficient inverse graphics in biological face processing. *Science advances*, 6(10), 2020.

[16] Hakan Yilmaz, Gargi Singh, Bernhard Egger, Josh Tenenbaum, and Ilker Yildirim. Seeing in the dark: Testing deep neural network and analysis-by-synthesis accounts of 3d shape perception with highly degraded images. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43(43), 2021.