The philosophy and Science of illusion

Sagar B Dollin, 2020HCS7012
MSc Cognitive Science,
Indian Institute of Technology, Delhi

To start, I would like to define an illusion. In my opinion, an illusion is something that we perceive, which isn't the ground truth. For example, if I see a magician taking out a penny from my ears, it is an illusion, and it is not true that there are pennies in my ears, which only magicians know how to take out. Also, before proceeding when I say *perceive* in this paper, I mean visual Perception. But what is Perception? In layman's terms, visual stimulus perception means the things we visually see or experience if one wants to add phenomenology to its definition. So at the most basic level, visual Perception can be considered as the experience of seeing things. One can see real and nonreal things. Nonreal things like things you see in your dreams. Can we say from our definition of illusion that the things we see in our dreams are illusions? Well, perhaps no. The thing is that dream is considered a separate phenomenon. Even if it is an experience or not[1], irrespective of that, I would not view dreams as illusions in this discussion at least. The reason being dreams occur only in our minds, and they are a reflection of our thoughts and the things we see. They aren't nonreal things but a schematic reflection of our own real experience in the world.

Since I have used the word real for a few times here, I think it is my responsibility to discuss it and what I mean. First of all, there could be debates that argue things we see aren't real; they are only representations of the real objects; hence, how can representations be real? Let me elaborate. When we see a red Square, we see a 4-sided object with approximately equal sides, and the light reflecting off the object is some spectrum of light that our eyes and brain interpret as Red. So to think of it, there is no red color existing in reality, but only the representation corresponding to some spectrum of visible light. So do we conclude that everything we see isn't real? That would be a regressive move to do so. Even though we know everything we see as a representation, we will say that the things that can be verified by different views under the same conditions as real and the things that can be verified as nonreal are illusions in real-time situations. For example, if I see a white puppy in a park, I can confirm that it is indeed a white puppy and not a black cobra. Hence it must be real. On the other hand, if I see the face of a white puppy in the clouds, I can say that it is an illusion because puppies cannot fly. I can confirm that if it were in clouds, it would be so distant from me that I won't be able to see it so clearly. It must be the clouds that resemble a white puppy. Another view that questions our view of real and nonreal visual Perception of things is the one that assumes that all the reality or experiences we perceive could be a part of a simulation. Is it possible that we are just a brain and are immersed in some liquid and all the experiences we perceive are the signals sent by wires connected to our brains? As attractive as the idea sounds, we would not consider this notion for our discussions since I am not Elon Musk. And also, for valid reasons, any cognitive scientist wouldn't conduct any research considering this as one possibility because we strive for parsimonious explanations of complex phenomena we witness in this field. And such considerations will only contribute to hindrances.

[1] Are Dreams Experiences? Author(s): Daniel C. Dennett

Problem Statement

Now that we are clear of any possible questions that can be posed on the assumptions and grouds we are considering in our discussions, we are ready to move on to the problem statement that I want to address. The problem with Philosophers of Mind and Cognitive Scientists is that they sometimes define the terms and language of their subject about one topic and diverge from each other, creating a long gap. Comparing both becomes very difficult because they are so different in terms of progress and discussions. Dennett also pointed this out in his paper "Towards a Cognitive Theory of Consciousness." Of how cognitive scientists have ignored the theory of consciousness and left it only for philosophers to debate. Hence Denett proposed this paper[2] to draw a connection between the two. I will be doing more or less the same in this discussion; I will draw inferences from Cognitive studies of illusion and then compare them and account for the philosophical concepts that I am aware of.

Case Study – Kaizsa illusion

Let us first take an example of a Kanizsa Illusion. The figure you see below is a Kanizsa square. But if you observe, you can verify that the only figures in this image are the Pacman figures, and they are oriented such that they create an illusion of square. We see the right angles perfectly opposite each other, and our brains are fine-tuned to see this as square. We even tend to perceive the edges that aren't there. The blank space between two consecutive Pacman has no edge for the square, but our brain suggests they are there. We even perceive the white square to be brighter than the white background. This is because our brains are fine-tuned to do a figure-ground separation. I will talk about both these terms(suggestions from the brain and figure-ground separation) mentioned briefly.

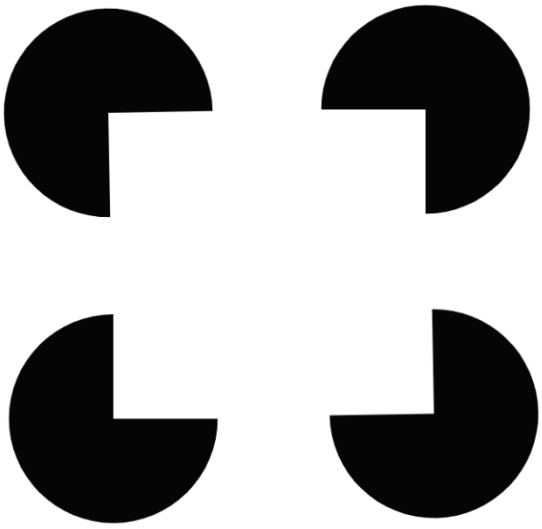


Image 1: Kanizsa Square

[2] Toward a Cognitive Theory of Consciousness- Daniel C Dennet

The Perception of edges of the square in the space between the Pacman's is nothing but a suggestion from our brain. I say so with confidence because it can be verified that there are no edges in that empty space by using a pixel value reader, and we can confirm that there are no pixels for edges in that area, and it is continuous. Now that we know that this is indeed an illusion, on what grounds can I say it is the brain that is suggesting. It could be anything else, or maybe nothing is being suggested, and this illusion has some other explanation.

Visual Perception Mechanism – A functionalist view

Before explaining why the idea of suggestions must be accurate, I would like to give some idea about the visual processing of the human brain. The inputs we receive in the eyes are transduced to electrical signals and sent to the visual area, occipital lobe present in the posterior region of the cortex of the brain. This occipital lobe can be seen to be divided into layers of neural networks hierarchically arranged. Example V1, V2..V5. Here V5 can be considered as a higher layer and V1 as a lower layer. Lower layers are good at detecting features like edges, corners, etc. The information flows from lower layers to the higher layers detect features that are a combination of features detected by the lower layers. For example, if the lower layers are detecting edges, the higher layers will put together those edges to detect the shapes formed by those edges. It is observed that information can flow in both directions, from lower layers to higher layers(bottom-up) and vise versa(top-down). But of course, there is no reverse transduction that is converting electric signals back to light waves. Once the representation is in the electrical signal, no matter the flow of information, the representations remain in the same electrical form.

To summarize, the higher layers perceive the global Perception, for example, identifying a square or any shape or object. On the other hand, even lower layers contribute to this Perception by processing smaller details, but we don't have access to those processed outputs of lower layers. We only perceive the bigger picture, we can zoom in the picture, but we cannot visually access the lower layer outputs.

It has been seen that there are predictive neural networks in our brain. In our visual layers of hierarchically stacked neural layers, the higher layers try to predict the outputs of the lower layer neurons and update the activity by using the error produced by this prediction. To simplify it, the layers in the V2 Layer of the visual area predict the output of the V1 Layer. And the error here is the difference between the predicted activity (*p*) and actual activity(*e*) of V1. This difference is used to update the activity of the neurons.

$$\text{Error} = p - e$$

Rao and Ballard first proposed this hypothesis[3], "in a hierarchical system, each layer tries to predict the activity of the layer below, and the prediction errors are used to update the activations."[4] But Rao and Ballard never mentioned suggestions or expectations. It was further researches where it was found that illusions can be

[3] Rao, Rajesh & Ballard, Dana. (1999). Predictive Coding in the Visual Cortex: a Functional Interpretation of Some Extra-classical Receptive-field Effects. Nature neuroscience. 2. 79-87. 10.1038/4580.

[4] Predictive coding feedback results in perceived illusory contours in a recurrent neural network Author(s): Zhaoyang Panga , Callum Biggs O'Maya , Bhavin Choksia , Rufin VanRullena,b,

perceived only if there are predictive coding networks like these in a computer simulation of visual recognition (Panga et al.) We will talk more about this computational modeling of predictive coding networks in the next section. I want to highlight here that a neural network model is capable of expectation if the higher layers (abstract levels) try to predict the lower layer activities (layers responsible for initial processing of visual information). When they predict, they are expecting the lower layer outputs to be a certain way. This expectation leads us to expect a square when we are shown Pacman figures that are oriented such that we see the corners of a square and we expect the entire square to be perceived. The lower layers in our visual system detect these corners and send the information in a bottom-up fashion in the hierarchy to the layers above. Since the corner appears to be that of a square familiar to our memory, the higher layers expect it to be a complete square in the image. So they predict that the lower Layer is trying to send a square, but in reality, lower layers are not sending a square. This leads to an error in prediction. This prediction error is used to update the neuronal activity between the two layers. This is the suggestion that I mentioned earlier. These errors being used to update the lower Layer's activity can be seen as the higher layers suggesting to the lower layers that it is a square. Hence we see the illusion of a square even though it is not an actual square with complete edges.

To simplify the above explanation, let's say the higher layers and lower layers could talk to each other. So when they get input from the eye, this is how their conversation might look like:

while(input is Kanizsa_square):

Lower Layer (L.L.): I detected the corners of a square.

Higher Layer(H.L.): Then, I expect there to be a complete square.

L.L.: I don't see the complete square. There are missing edges.

H.L.: There is an error. I suggest you update your activity and perceive this as a complete square. There must be some missing information in the stimuli.

L.L.: Yes, I can perceive the missing edges with the suggestions.

This might not be the aptest explanation, but it is good enough to explain how a predictive coding network can suggest perceiving an illusion like Kanizsa square.

I have two questions at this point. Why would we have such a network? And how can we empirically verify that such a network can indeed lead to suggestions that sometimes lead to perceiving illusion?

Answering the first question is pretty simple but can again lead to various other questions that are not our scope for now. We are always looking at partial i=visual information. We can never look at a complete 3D object at once. There is always some part hidden. But we know how that object would look like even though we don't have

visual access to all its details. Our visual network is fine-tuned to deal with partial information. We use the mechanism of expectation and suggestions to deal with everyday visual Perception. If we see only the edges of a laptop on the table, we know that the whole laptop is there. We are fine-tuned to expect complete objects from partial information. Also, when we see faces in the clouds, there aren't actual faces, but since some features appear like faces, our visual networks are fine-tuned to suggest that they are faces. This pattern recognition is maybe an evolutionary advantage. We are good at looking at patterns and perceive them as complete patterns. I will not debate whether or not animals have similar mechanisms, but I will say this kind of mechanism is advantageous. It might have helped our ancestors survive to perceive a deer that is camouflaged in the dry grass. Or look at a stone-age tool and suggest what is missing in the tool to make it more useful.

Anyway, without digressing further, I would like to give a small account of Daniel Denette on suggestions and expectations. Daniel Dennett speaks about a painting '*Delloto: View of Dresceden*' where he sees several people on a bridge, and he wonders about what kind of clothes they are wearing and how they look like[5]. But when he zooms in the painting, he realizes that these people on the bridge do not have as many details as he imagined. Dennett points out that his brain has taken a suggestion because it is expecting people to have colorful clothes and hands and legs. Still, the painting did not have so many details as the people looked tiny and distant. So the well-known philosopher who has good cognitive knowledge would agree to our mechanism of suggestions. He may or may not acknowledge the prediction coding network, but the idea of suggestions and expectations resembles both the accounts. Also, to remind you that I'm not suggesting any ideas as my own, I'm only drawing a connection between cognitive research of illusion perception and philosophical views on the same.

The second question of how this connection of predictive coding networks can lead to the described expectations and suggestions be verified? Number one: We can do a neuroscientific study and see how the neurons interact when presented with illusory contours. If we were to observe all the neurons and their interaction while this happens, we could compare the results with our hypotheses. As simple as it sounds, it could be challenging to achieve for all the reasons you can think. So not practical at this time, but attempts have been made to compare computational models with the visual cortex and give promising similarity in the learning procedure[6]. We can build a neural network model and test if it perceives the kanizsa illusion, this may not provide the hard evidence, but it will help us evaluate the hypotheses. I will answer the questions surrounding the idea of the computational model using to evaluate suggested theories in the philosophical account at the end.

But before moving to the next section of computational models, let me discuss the figure-ground Perception that I left open in the beginning.

---

[5] "Daniel Dennett: Consciousness Explained (2013 WORLD.MINDS ...." 31 Dec. 2013, https://www.youtube.com/watch?v=JP1nmExfgpg.

[6] R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, A. Oliva, Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence, Scientific reports 6 (2016) 27755.

Figure-ground Perception is nothing but separating the figure(object) from the ground(background). Once we have done that, we start perceiving depth. This allows us to see the figure closer to us than the ground. The same happens in the kanizsa square as well. Since the square and background are white, we perceive the square to be brighter than the ground, making it appear that the square is floating on top of the ground; hence it appears closer. Mark this point because this will help us answer philosophical questions like does a computational model perceive images?

Computational models for illusion perception

The most common computer vision models use CNNs. They are Convolutional Neural networks stacked on each other, hierarchically. Still, the flow of information is unidirectional, i.e., only bottom-up. Hence they do not have predictive networks which require the flow of information bidirectional.

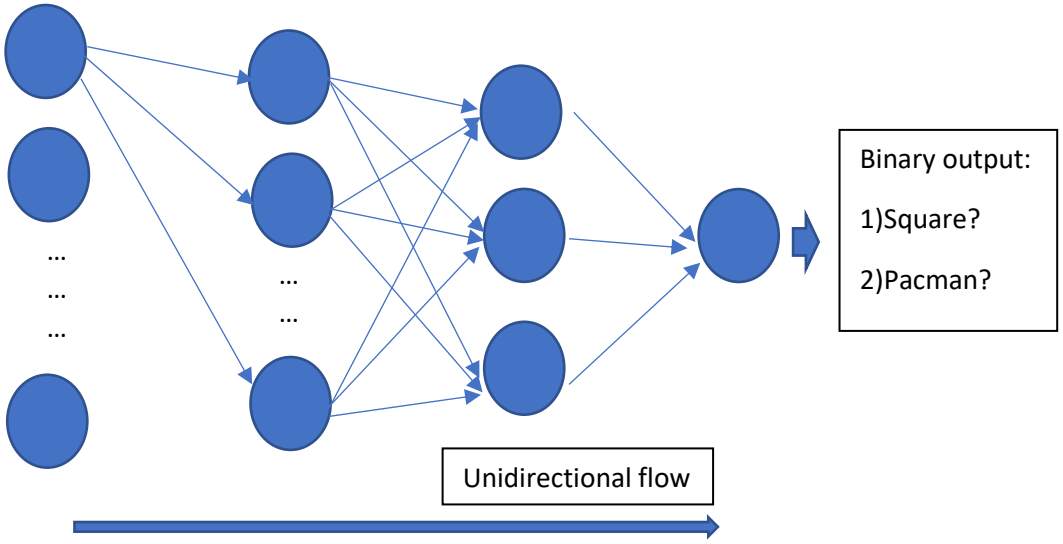


Image 2: A standard CNN to recognize if an input image is a square or pacman

It has been observed that when this kind of CNN is trained on real square images and random orientations of Pacman to identify them as square or Pacman, and when tested on Kanizsa square, it recognizes it as Pacman and not as square. This is because there are no suggestions or expectations involved here.

Let us have a look at the predictive coding neural network defined by Zhaoyang et al. in the paper; *Predictive coding feedback results in perceived illusory contours in a recurrent neural network*.

The model consists of 3 feedforward encoding layers, e1,e2,e3, and three corresponding decoding generative layers d0,d1,d2, whose errors are used to update the activity of the encoding layer at each timestamp. The error of a layer n at any timestamp t can be given as,

$$\epsilon_n = e_n - d_n.$$

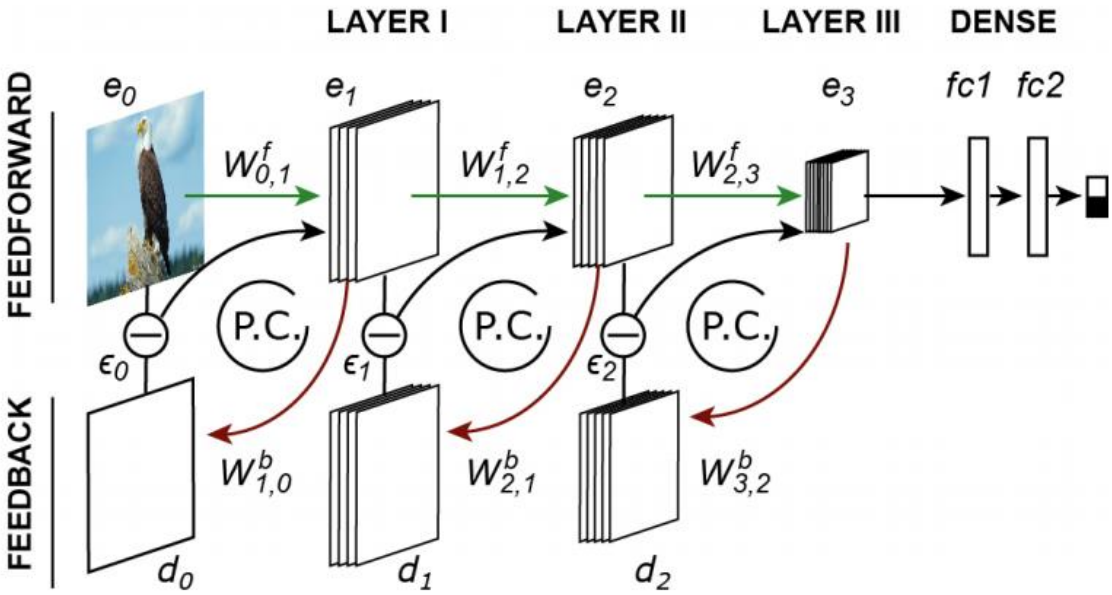


Image source: *Predictive coding feedback results in perceived illusory contours in a recurrent neural network*

Image 3: e1,e2,e3 are the feedforward encoding layers whose activities are updated(P.C loops) using the errors($\in_n$) produced by the corresponding d0,d1,d2 feedback generative layers that try to predict the activity of the higher e_n layers.

To the hierarchical encoding layer, we attach a binary classifier. That is used to classify images as either square or Pacman.

The encoder model with the predictive coding network is initially trained on the Cifar100 dataset of natural images(tuning). It has been found that it is essential for neural networks to be trained on natural images to develop expectations for recognizing patterns. Then we add the classification head to the encoding model and train the complete model on square images and random Pacman orientation images, as shown below.

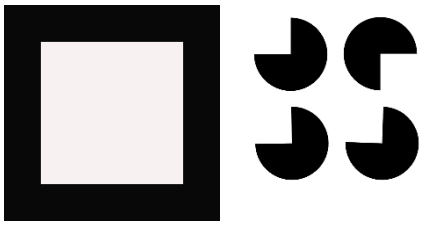


Image 4: (a) shows the square image and (b) shows Pacman in random orientations.

Once the model is trained on square and random Pacman orientation, we test our model on the Kanizsa square and Pacman orientation so that all are facing up, as shown below.

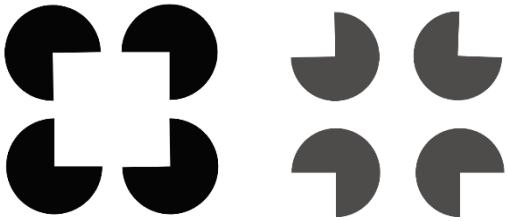


Image 5:(a) shows the kanizsa square image and (b) shows Pacman in outward orientations.

We can intuitively see how this works; the model is initially trained on real square images and Pacman in random orientations. It is then tested whether it identifies the Kanizsa illusion as a square or as Pacman. And for all outward Pacman orientations, it is expected to identify it as Pacman. Hence our classifier model had a binary classifying head; the image is classified as either square or Pacman.

This kind of model identifies the Kanizsa illusion as square and not just as Pacman as in the case of CNN models. Please note that the model I described is not my own, but I provided this model from the reference Zhaoyang et al..

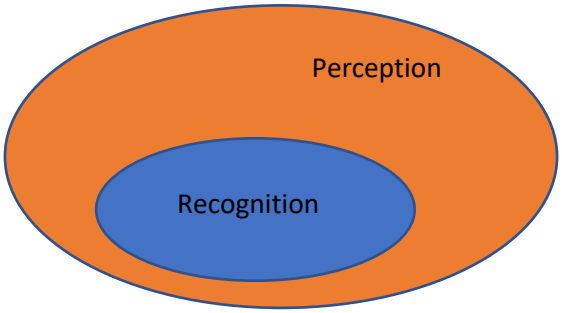
Philosophical Account

When we talk about the computational model, the first thing that could strike as a question is how computers can recognize some images compared with Perception? The author of the paper of this computational model answers this question on the surface using technical methods.

The author addresses that even though the network classifies a kanizsa square as a square with a higher probability than that of CNN models, does it mean that the model actually "sees" the illusion? A predictive coding feedback model is also a generative model, meaning the same image can be generated given an input. And when tested on Knaizsa square, the image generated by the model had a square brighter than the white background. This result throws us right back to the discussion of the figure-ground Perception. In the rendered image, the model is trying to separate the square as a figure from the ground by making it brighter, and hence the author concludes that this is comparable to actual Perception.

But our discussion will not end here. In this entire setup, we might have used or understood recognition and Perception interchangeably. So does it mean they are both the same? One would scream to say NOO! But we will calmly discuss why they are different yet could be used interchangeably in some instances like the above. Let's define what Perception is. One could say it is a phenomenological experience. Still, I assume Daniel Dennet would argue since, in his work towards a cognitive theory of consciousness, he systematically shows why the experience of consciousness itself is an illusion of self. One has no access to the processes going on in our head but only to the outputs of some accessible processes and appears as an illusion of self. In this view, one can also define Perception in terms of functionalist. Perception can be defined as the response of our visual mechanism to the visual inputs. The outputs of these visual systems that we have access to become the Perception that we

experience. So to answer Perception, the question breaks down to what are things we have access to. We can say we have access to the memory of objects, hence object recognition is a part of Perception. We also estimate how far the object is, not actual metric distance but an approximation. Therefore we can say depth is also a part of Perception. And many other things like the orientation of objects, the velocity with which they or we are moving relative to each other, and other things that I may have missed.



If you see the above diagram, it shows the relationship between Perception and recognition through the Venn diagram. It can be seen that Perception and recognition can be interchangeably used only when we are explicitly talking about Perception involving only recognition, as in the case of Kanizsa illusion, it is only the Perception of recognizing the illusion and not other aspects. Of course, we also talked about depth perception in figure group separation, but the heart of the problem remains to recognize the illusion.

But another question remains open: Can a computer have Perception because it is nothing to be like that computer? Dennet already answers the answer to this. It is something to be like you because you have access to outputs to ceratin processes that creates this illusion of being you. Similarly, if we make a hierarchical system with abstracting access levels, we could, in theory, simulate something that believes it is something to be that thing. Still, in reality, it is like nothing to be that thing. So we can say that as long as a computer behaves as we do to the visual simulations, it is experiencing an illusion of being something in the world. Or at least it is creating an illusion to us as it is something to be like it.

References

1) Are Dreams Experiences? Author(s): Daniel C. Dennett
2) Toward a Cognitive Theory of Consciousness- Daniel C Dennet
3) Rao, Rajesh & Ballard, Dana. (1999). Predictive Coding in the Visual Cortex: a Functional Interpretation of Some Extra-classical Receptive-field Effects. Nature neuroscience. 2. 79-87. 10.1038/4580.
4) Predictive coding feedback results in perceived illusory contours in a recurrent neural network; author (s): Zhaoyang Panga , Callum Biggs O'Maya, Bhavin Choksia, Rufin VanRullena
5) "Daniel Dennett: Consciousness Explained (2013 WORLD.MINDS ...." 31 Dec. 2013, https://www.youtube.com/watch?v=JP1nmExfgpg

6) R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, A. Oliva, Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence, Scientific reports 6 (2016) 27755