Do we have the unified theory of perception?

Our previous experiences, according to Hermann von Helmholtz, aid us in deducing the cause from the effect. To express this notion, he concentrated on the domain of visual perception (Yantis, 2001). This indicates that the brain tries to infer the causes based on sensory input, resulting in a conclusive action. This idea has been in extensive research over the last few decades due to the availability of mathematical frameworks to support it. The Bayes theorem of probability has emerged as a prominent framework that supports the process of cognitive inference (Thagard, 2019; Rescorla, 2020)

Thomas Bayes initiated a theory of finding conditional probabilities based on observations of the world. There are three main elements in this theorem, the "prior", which is some previously known attribute about the world before making an observation. The "likelihood" is an estimate of observing something in a given situation and the "posterior" is an update to the prior by combining it with the likelihood which is the newly gathered information (McNamara et al., 2006). For example, consider an office where the cleaners come every Wednesday. The prior would be the cleaners coming every Wednesday as this information is previously known so the posterior would be the same as prior without any new observation. It rains one Wednesday, and the cleaners do not show up. The posterior needs to be updated to account for this new information that the cleaners do not come on Wednesdays when it rains.

The ability of the brain to form representations for statistical correlations is demonstrated with an example by Helmholtz. In his example, he suggests that if there are two events naturally occurring together such as thunder and lightning, the brain can deduce a commonality between them. Such deductions can help in inference and future predictions. However, he also points out that simply relying on a single observation without experimentation could be erroneous as we might not be able to account for all the factors that lead to the result i.e. thunder and lightning occurring together in this case (Yantis, 2001; Thagard, 2019). This concern is corroborated by the fact that observations as a result of sensory input can be interpreted in multiple ways. For example, the shadow of two threedimensional objects could look the same, words like "eye" and "I" sound the same, and multiple objects could smell, taste, and feel the same. This makes it difficult to ascertain the interpretation of the world despite the sensory signals being high in resolution. Furthermore, the sensory input can be noisy and low-resolution, for example, an object that is kept in a dark room or incoming traffic in foggy conditions, which can cause additional errors. Due to this, the brain can only provide its best estimation of the surroundings through inference (Ji Ma et al., 2012).

The use of the Bayes theorem can be explained in the context of cognitive science. The prior would remain the same i.e. having some previous knowledge about the situation. The likelihood would be a probability estimation of the sensory inputs due to observation and the posterior is the final prediction that integrates all the information from prior and the likelihood. It is important to note that the sensory input is continuous; for example, vision corresponds to distance and light intensity, while sound corresponds to volume, which is why the outputs are represented as probability density curves rather than discrete probability values in this context (Baykova, 2022; Baykova, 2021). The process of using the Bayes theorem as the mechanism for inference is called Bayesian inference.

The Bayesian inference model of perception has been shown to successfully fit various empirical experiments (Rescorla, 2020). To explain the physical implementation of perceptual inference in the brain, an extension to the Bayesian inference model was proposed called "predictive processing". The concept of predictive coding originated as a result of experimentation in the visual system. Srinivasan et. al. (1982) suggested that the lateral inhibitory response in the visual system follows a similar process in video transmission where the signal amplitude is reduced by removing redundant and predictable components. In their predictive coding model, the intensity of the surrounding cells is used to generate an estimate of the incoming intensity at the principle cell. This estimate is subtracted by the incoming signal to reduce the amplitude of the signal which helps in separating the valid signal from the noise (Srinivasa et al., 1982).

Rao and Ballard (1999) postulated a more elaborate model of predictive processing using the hierarchical structure of vision. In their model, the neuron at the upper level sends a prediction about its activity to the lower level neuron. The lower level neuron compares this prediction with the incoming signal from the neuron below it to calculate the error. This error is the signal that is passed to the upper-level neuron which was the "incoming" signal. Rao and Ballard simulated the property of extraclassical surround in the receptive field with their model of predictive processing. This property inhibits the response if a stimulus is extended from the center to the surrounding area. In terms of predictive processing, the presence of any stimulus in the surrounding area can be used to predict its presence in the center. Hence, the difference in the incoming signal at the center and the estimated probability would be low therefore the error and the firing rate would be less (Rao & Ballard, 1999; Lindsay, 2021). It is to be noted "predictive processing" is the umbrella term for "predictive coding".

In general, the predictive processing model represents the brain as a layered structure that interprets the sensory information using approximation of Bayesian inference in each layer to generate predictions and minimize errors over a long run. The use of Bayesian inference is not exact. The prediction errors are passed to the upper levels in feedforward connections and the prediction of the current level is passed down to the lower level in a feedback connection. This prediction from the upper level is used to compute a posterior by multiplying it with the error signal from the lower level (Baykova, 2022). The focus of this essay is to discuss if predictive processing can account for human perception and decision making.

Predictive processing is a generative model as it continuously generates predictions about the sensory input at multiple levels. If there is some unexpected input, the prediction error will be high. (Wacongne et al., 2011) test this paradigm with an auditory mismatch negativity test (MMN). They selected ten healthy subjects for this experiment with no known neurological issues. The audio setup consisted of sound waves of two distinct superimposed sinewave sounds A and B. The experimental setup was a permutation of blocks presenting 1) five identical sound waves (AAAAA)/(BBBBB). 2) four identical sound waves and one different wave (AAAAB)/(BBBBA). 3) just four identical sound waves (AAAA)/(BBBB). Each block started with a repetition of 25 frequent sounds to establish the global expectation. For example, in block 1 (AAAAA) or (BBBBB) was presented 25 times to establish that this pattern of the auditory signal should be expected. After setting up the global rule, 100 random occurrences for the same block were presented of which 75 were the same as the global rule i.e. (AAAAA)/(BBBBB), 15 were from the deviant group i.e. (AAAAB)/(BBBBA) and the last 10 were from the omitted group i.e. (AAAA)/(BBBB). Similarly, block 2 sets the global rule for the deviant group. And block 3 sets the global rule for the omission group with the exception that all the 100 occurrences were identical to the global rule. Simultaneous EEG and MEG signals were recorded. The subjects were asked to pay attention to the sound and fixate on a cross to avoid eye movements. The goal was to determine 1) whether an unexpected sound generates a larger response. 2) whether some "expected" prediction is generated when a sound is omitted from a sequence. 3) whether learning the sequence from the deviant group reverses the result. 4) Since the brain learns the unexpected sound, its omission should create a higher prediction error when generating an estimation. The authors show that the results are consistent with the concept of predictive processing. The unexpected sound induced in blocks 1 and 2 showed a high MMN response. The global rule for the deviant block was learned a high MMN response was generated for five identical tones. Finally, an early omission effect was generated just after the fourth tone when a sample from the omission group was presented. This corresponds to the generation of an “expected” prediction i.e. the expected sound.

The experiments proposed by (Todorovic et al., 2011) are also consistent with the predictive processing model and supports the notion that expected predictions generated using the top-down hierarchy in the predictive processing model inhibits the response. In their experiments, the subject listened to an auditory stimulus in blocks where the sound repetitions were expected or unexpected. They recorded the neural activity of the subjects using MEG and arrived at the conclusion that the predictions generated through expectations, inhibits repetition response.

From (Wacongne et al., 2011 and Todorovic et al., 2011), it is implied that predictive processing inhibits the response if an expected input is received and increases the response if an unexpected input is received. However, (Kok et al., 2012) show that this behaviour of predictive processing can also change depending on the task. They proposed an attentional cueing experiment that enhances the predicted response rather than inhibiting it. The authors suggest that predictions based on attention increases the response to improve the precision of perception. 22 subjects were selected with normal or corrected to normal vision for a grating orientation identification task out of which 5 were removed due to excessive head movements and the inability to understand the task. The experimental setup comprised of a prediction cue block and an attention cue. A prediction cue was presented at the start of each block that stated the likelihood of the position of the stimuli. For example, (“In the following block of trials, stimuli will likely appear on the right”) and the same for the left. A neutral prediction cue was also present that did not specify any direction. Each block is followed by 8 trials where attention cues are shown at the start. The attention cue contained a small triangle that indicated which half of the attention cue the subjects needed to pay attention to. The subjects were instructed to identify the orientation of the stimulus only if it appeared in the half where the attention cue triangle indicated. This essentially compelled the subjects to pay attention as the prediction cue was 75% accurate in indicating where the stimulus might appear. Further, the triangle indicating the attention area in the attention cue is not necessarily the same as indicated by the prediction cue. For example, the prediction cue might state “In the following block of trials, stimuli will likely appear on the right” with 75% certainty but the triangle in the attention cue might be pointing left. This means that the subject had to pay attention to the stimulus appearing in the left part of the attention cue with a probability of 25%. The stimulus appearing on the opposite side could be ignored and no response was required for it. They proposed two hypotheses, one where the prediction response and attention response have opposing effects i.e. prediction is inhibitory as per the “norm” and attention is excitatory that outweighs the prediction response and a high response is generated for predicted and attended stimuli. The other is where both prediction and attention work in synergy to increase the precision of predictions leading to an increased response for predicted and attended stimuli. The majority of the subjects successfully paid attention and identified the gratings for the stimulus that appeared on the relevant side. The fMRI results indicated that prediction and attention together increased the precision of prediction validating hypothesis 2 and confirming that instead of inhibiting the prediction response, attentional cueing can reverse this behaviour.

Other studies such as (Arnal et al., 2011) experimented with disparity between multisensory inputs such as vision and audio. They concluded that the results are consistent with the predictive processing model of the brain wherein bottom-up prediction errors are transmitted to the upper levels as high frequency gamma waves and top-down predictions are transmitted via slower beta waves.

Does this mean that predictive processing can robustly account for human perception and decision making? The answer seems to be a bit complicated. While all the experiments mentioned above does result in conclusive evidence that the theoretical framework of predictive processing can explain biological perception and decision making, there is a lot about predictive processing that is ambiguous. There is a specific focus on (Wacongne et al., 2011) and (Kok et al., 2012) to discuss this ambiguity. If multiple studies agree that the top-down prediction response is inhibitory to minimize the error, the fact that attention can reverse this inhibitory behaviour (Kok et al., 2012) makes this model a bit fragile because it means that there might be many mechanisms such as attention that can change the "behaviour" of the predictive processing framework thereby making it unpredictable. Also, there are studies that can account for the same behaviour without the use of predictive processing at its core. The extraclassical suppression postulated by Rao and Ballard can be explained without the use of predictive processing as shown by (Coen-Cagli et al., 2015). They show that the surround suppression is dependent on different sensory inputs. They also highlight the importance of first understanding possible cortical responses based on stimulus instead of deriving descriptive models.

Furthermore, critical reviews of the predictive processing framework points to the challenges and possible reasons for neglecting alternate theories. One very noticeable was the lack of discussion about any alternate approach with their advantages or disadvantages in studies that show consistency with predictive processing. (Colombo et al., 2021) argues that the Bayesian approach due to its applications in multiple fields is researched more frequently by prominent scientists. Hence, a variety of tools are available to use from interdisciplinary fields which is why the wide adoption is not necessarily due to novel research. (Williams, 2018) on the other hand contends that predictive processing does not account for the generality and the rich composition of thought. He states that humans are capable of abstract thought and reasoning at any spatiotemporal scale without necessarily taking in sensory input and predictive processing cannot combine representations across such levels. Also, the characteristics of thought are compositional in nature, meaning, a thought displays a systematic structure. Since there are multiple domains a thought could occur in for example, language, a recursive model like predictive processing could result in infinitely long compositional representations. The core issue of this model as Williams suggests is its lack of ability to account for the productive combination of flexible and systematic concepts that produce a thought.

To summarize what has been discussed so far, Helmholtz's idea of "unconscious inference" inspired research along the lines of thinking about brain as a prediction engine. The brain is able to provide a best estimation as the sensory information can be interpreted in multiple ways which required a framework to study how the brain might be predicting the sensory information. Due to its popularity, the Bayes theorem of probability emerged as the prime candidate framework to study and explain the biological process of predictive inference. With this framework, the idea of Bayesian brain emerged that extended into predictive processing. Predictive processing represents the brain as a layered structure to transfer prediction errors in a bottom-up way and generate predictions in a top-down way to minimize error. This model agreed with multiple empirical and neural evidence that explained how the brain might be minimizing errors by inhibiting expected input and aiding in increasing the precision of prediction in one case. However, the performance of this model in the context of biologically explaining human perception and decision making is weak. As there are discrepancies in the behaviour of the model when applied to different mechanisms of perception, a question arises if there are other mechanisms that might change the output of predictive processing. Also, the phenomena that can be explained by using this model have been shown to provide similar results without it. Critical reviews about predictive processing points to concerns that cannot be overlooked. They discuss about fundamental issues of negligence towards alternative approaches and why this model cannot generalize well due to its architecture limitations. The fact that there are a lot of disagreements, explanation gaps, and ambiguity demonstrates the need to do further research in predictive processing. In its current situation, answer to the question that human perception and decision making is well accounted for by predictive processing or if we have a unified theory of perception remains ambiguous as well.

References:

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–801. https://doi.org/10.1038/nn.2810

Bayesian behaviour, predictive brains: Neuroscience of Consciousness [21/22]. (2022). Sussex.ac.uk. https://canvas.sussex.ac.uk/courses/17225/pages/lecture-6-bayesian-behaviour-predictive-brains

Baykova, R. (2021, July 1). *Investigating the role of Bayesian inference in duration perception*. Sro.sussex.ac.uk. http://sro.sussex.ac.uk/id/eprint/100003

Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience*, *18*(11), 1648–1655. https://doi.org/10.1038/nn.4128

Colombo, M., Elkin, L., & Hartmann, S. (2021). Being Realist about Bayes, and the Predictive Processing Theory of Mind. *The British Journal for the Philosophy of Science*, *72*(1), 185–220. https://doi.org/10.1093/bjps/axy059

Ji Ma, W., Paul Kording, K., & Goldreich, D. (2012, May 5). *Bayesian models of perception and action*. Www.cns.nyu.edu. https://www.cns.nyu.edu/malab/bayesianbook.html

Lindsay, G. (2021, June 3). *The Challenges of Proving Predictive Coding*. Www.simonsfoundation.org. https://www.simonsfoundation.org/2021/06/03/the-challenges-of-proving-predictivecoding/#:~:text=One%20of%20the%20earliest%20theories

McNamara, J. M., Green, R. F., & Olsson, O. (2006). Bayes' theorem and its applications in animal behaviour. *Oikos*, *112*(2), 243–251. https://doi.org/10.1111/j.0030-1299.2006.14228.x

Peter Kok, Dobromir Rahnev, Janneke F. M. Jehee, Hakwan C. Lau, Floris P. de Lange, Attention Reverses the Effect of Prediction in Silencing Sensory Signals, *Cerebral Cortex*, Volume 22, Issue 9, September 2012, Pages 2197–2206.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rescorla, M. (2020). Bayesian modeling of the mind: From norms to neurons. *WIREs Cognitive Science*. https://doi.org/10.1002/wcs.1540

Srinivasa, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *216*(1205), 427–459. https://doi.org/10.1098/rspb.1982.0085

Thagard, P. (2019). *Cognitive Science (Stanford Encyclopedia of Philosophy/Summer 2019 Edition)*. Usyd.edu.au. https://stanford.library.usyd.edu.au/archives/sum2019/entries/cognitive-science/

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. *Journal of Neuroscience*, *31*(25), 9118–9123. https://doi.org/10.1523/jneurosci.1425-11.2011

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, *108*(51), 20754–20759. https://doi.org/10.1073/pnas.1117807108

Williams, D. (2018). Predictive coding and thought. *Synthese*, *197*(4), 1749–1775. https://doi.org/10.1007/s11229-018-1768-x

Yantis, S. (2001). Visual Perception: Essential Readings. In *Google Books*. Psychology Press. https://books.google.co.uk/books?hl=en&lr=&id=NzzaI2lCsN4C&oi=fnd&pg=PA24&dq=von+Helmholtz+H.+3rd+edn.+Vol.+III.+1866.+Concerning+the+perceptions+in+general.&ots=MnadkPKiPo&sig=RLiG0FbnVTDVnfksjzEGAT2Gajc#v=onepage&q=infer&f=false