"I did not see that coming": Investigating predictive coding in the brain using EEG

Otto Sejrskild Santesson, (202004347@post.au.dk)

Faculty of Arts, Aarhus University

Bachelor of Cognitive Science

May 31st, 2022

This paper may be made public

Otto Sejrskild Santesson (202004347)

31st of May, 2022

202004347@post.au.dk

BSc Cognitive Science, Aarhus University

Abstract

Different accounts of the underlying operational principles of the brain have been put forth throughout decades of scientific research, although the majority have not stood the test of time and scientific inquiry. One of these early proposals includes the projection theory, which suggests that the underlying physiology of perception works by a stimulus causing its response (i.e. perception) directly. Hermann von Helmholtz' (1821–1894) research showed that this was not the case, which led him to propound an alternative theory coined the sign theory. According to his theory, perception symbolizes the stimuli from which they originate from, but are not one-to-one copies of them. Moreover, he argues that the brain makes serial mental adjustments based on learned experience to piece together a complete and coherent picture of the sensory information obtained from the environment (Patton, 2018). Since Helmholtz, scholars have used his theory of human perception as a stepping stone to formulate and test more intricate and comprehensive theories on how the brain incorporates previous experience to shape their actual perceptual experience. One of the most established theories that incorporates this notion as its main premise is predictive coding. Using predictive coding as the theoretical foundation, this pilot conducted an experimental setup adhering to the oddball paradigm in order to investigate how foregoing stimuli affects the prediction of the upcoming stimulus. To gauge this influence, the prediction-error measured by the amplitude of the P300 component for the following three different types of stimuli was determined: the standard, consisting of a sequence of five digits, a deviant where the second digit in the sequence deviated from the standard and a deviant where the fourth digit deviated from the standard. The amplitudes of the different stimuli's P300 component was compared using t-tests, which revealed that more congruent digits preceding the deviate did not elicit a stronger prediction-error, thus contradicting our hypothesis formed on the basis of predictive coding. Reasons to why we do not observe an otherwise robust effect are considered and suggestions of alterations to the experimental setup are made.

Keywords: Predictive coding, prediction-error, P300, EEG

Otto Sejrskild Santesson (202004347)

202004347@post.au.dk

31st of May, 2022

BSc Cognitive Science, Aarhus University

1. Introduction

1.1 Predictive coding

Predictive coding is a theoretical framework that offers an unified explanation of cortical function – namely that the core function of the brain is to minimize prediction errors, or in other words, surprise. This prediction-error is measured by comparing the incoming, bottom-up sensory signal, with the top-down predictions of what we expect the signal to be. If the bottom-up signal and the prediction do not match, it results in a prediction error, which is then utilized to update the internal models following a Bayesian modus operandi (K. Friston, 2005; Huang & Rao, 2011; Millidge et al., 2022).

The question might arise why the brain does not just process the signal as it is when updating its internal models in lieu of the prediction errors. The answer lies within the nature of the sensory environment that is the source of the bottom-up signals; given the nature of its immense complexity and variability, if perception only relied on an unbiased bottom-up processing, the amount of processing required will greatly surpass the capacity of the sensory receptors. In other words, simply providing a direct representation of the raw sensory data is not feasible for the perceptual system's capacity (K. Friston, 2005; Huang & Rao, 2011). Due to neurons being energetically costly, there is an incentive to optimize the use of the brain's limited resources for the betterment of the individual's evolutionary fitness. To hone the effectiveness of the information processing, the brain exploits the fact that the sensory environment is, to a great extent, redundant; meaning that a considerable amount of the incoming sensory information is both spatially and temporally correlated. To make use of this redundancy, it is therefore more economical to have internal models that generate top-down predictions of the sensory data and then update these models with via the prediction errors, if the bottom-up signals cannot be explained by the predictions (Clark, 2013; Huang & Rao, 2011; Millidge et al., 2022).

This form of redundancy reduction is analogous to efficient video processing: due to pixels tending to be spatially correlated with neighboring pixels, the intensity of a pixel can be predicted based on the surrounding intensities, thus "allowing the input to be efficiently coded as the residual error between the actual intensity and the prediction based on the surrounding pixels." (Huang & Rao, 2011, p. 581).

Additionally, due to pixels' tendency towards being positively correlated in close temporal proximity (i.e. sequential frames), it is possible to predict in the same manner as described above (Huang & Rao, 2011). The video processing therefore only needs information of which pixels to change the intensity of, meaning it only needs the deviants. Like with predictive coding in the brain, it is the input (i.e. sensory) data that deviates from the prediction, that provides the most information and is therefore prioritized based on the internal models. These models follow Bayesian principles and thus use the prediction errors between the signal and the prediction to update themselves in order to improve future predictions. If the signal does not deviate from the internal models' prediction, then it doesn't provide any new information that would change the *maximum a posteriori* (K. Friston, 2005, 2010).

According to predictive coding, these internal models are organized in a hierarchical structure, meaning that at each level a compromise is being made "[...] between bottom-up information about sensory inputs, provided by the level below and top-down predictions (or priors) provided by the level above." (Garrido et al., 2009, p. 4). The prediction and error-correction loops take place simultaneously throughout the hierarchical levels, thus forming a sequential line of Bayesian-like model updating.
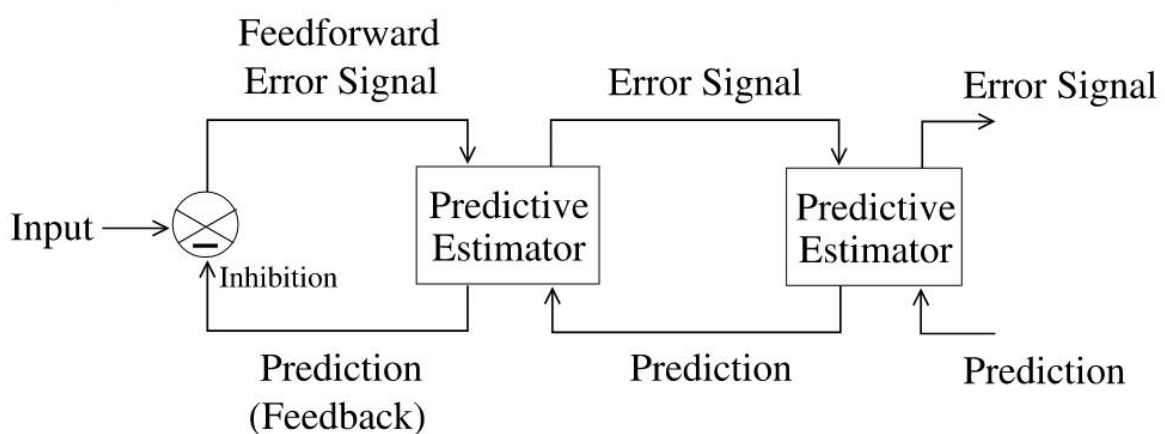

Figure 2: *The general structure of a hierarchical predictive coding model. At each of the levels, feedback pathways send predictions of the bottom-up signal at the lower level, while feedforward pathways send the prediction-errors between the predictions and the actual bottom-up signal. The errors are used to update the model, which then forms a new prediction (Figure reprinted from Rao & Ballard, 1999).*

Otto Sejrskild Santesson (202004347)

202004347@post.au.dk

31st of May, 2022

BSc Cognitive Science, Aarhus University

1.1.1 Predictive coding in the visual cortex

Evidence from neurobiological studies investigating the structure of the visual processing system indicate that it follows a structure that adheres to the predictive coding framework. Like with video processing, natural environments are similar in the way that image points that are close in space or in time tend to have strong positive correlation between intensities (Dong & Atick, 1995; Field, 1987; Hosoya et al., 2005). The visual system exploits this fact already in the neurons of the first stage processing, namely in the retina (Bear et al., 2020; Hosoya et al., 2005; Millidge et al., 2022; Srinivasan et al., 1982). Initially, the photoreceptors convert light energy into neural signals, where the intensity of the signals depend on the amplitude of the light waves. The intensities of these raw sensory signals are then fed through the bipolar cells into neural circuits in the retina consisting of ganglion cells, which predict the local intensity of the image point from neighboring intensity values. Sequentially, the predicted value is subtracted from the actual intensity of the image point (Bear et al., 2020). By performing this type of computational strategy, the ganglion cells do not signal the raw visual image, but moreso " [...] departures from the predictable structure, under the assumption of spatial and temporal uniformity" (Hosoya et al., 2005, p. 71). Similarly to the video processing example given above, it is the differences in illuminations of the ganglion cells' receptive fields that produces a response in the ganglion cells (Bear et al., 2020) (see Figure 2).
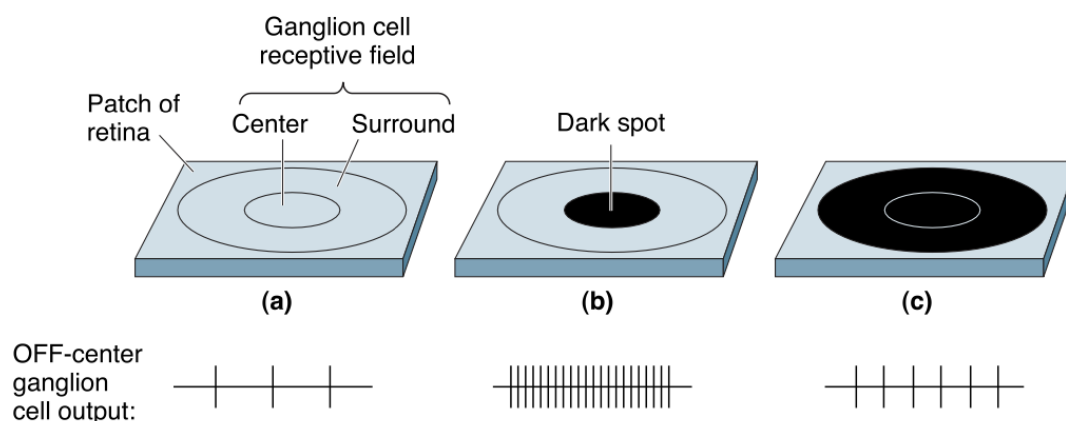
Figure 2: *An OFF-center (i.e. light in the center decreases activity whereas darkness increases activity) ganglion cell's response to different instances of receptive field; activation is decreased whenever the center matches the surround (a & c), while it increases when it is only the center that is darkened (b). Activity is decreased for both ON- and OFF-center ganglion cells if the center matches the surround, although if the center matches the polarization input requirement of the cell (i.e. dark*

for OFF cells, light for ON cells), the activation is less decreased (Bear et al., 2020) (Figure reprinted from Bear et al., 2020).

The signal, in the form of trains of action potentials, then propagates upstream through the optic nerve to a cluster of cells in the dorsal thalamus named the lateral geniculate nucleus (LGN). Whereas the spatial redundancy is mostly filtered in the retina, the temporal redundancy is believed to be removed in the LGN (Huang & Rao, 2011). From the LGN, the visual information feeds forward into the primary visual cortex via optic radiation (also referred to as *V1, area 17* or the *striate cortex*) and then onto the remainder of the visual cortex, V2, V3, V4 and V5 (see Figure 3) (Bear et al., 2020; Huang & Rao, 2011; Millidge et al., 2022). Along the way in the geniculocortical pathway, the visual sensory information is processed in increasingly higher order levels corresponding to different stimulus features, e.g. at lower levels, features such as whether the stimulus is vertical or horizontal are being processed, whereas at higher levels, more complex attributes are being processed, such as whether the stimulus resembles a hand or not (Bear et al., 2020; Millidge et al., 2022). The different areas of the visual cortex and the LGN are believed to interact with each other in terms of feedback connections from a higher level to a lower level (e.g. V2 to V1). For example, to provide predictions of expected activation patterns in V1, feedforward connections transmit to V2 the residual activity in V1 that deviated by V2's predictions (Rao & Ballard, 1999; Spratling, 2010).
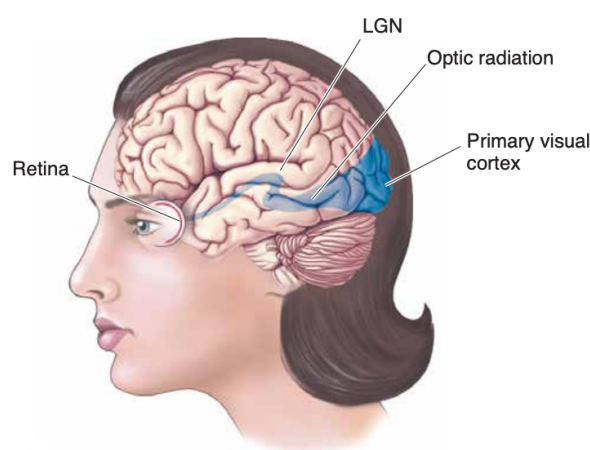

Figure 3: *A sagittal view of the brain with the retinogeniculocortical pathway highlighted in blue (Figure reprinted from Bear et al., 2020).*

The visual cortex is not the only perceptual system thought to be comprised of these hierarchical feedback loops – the predictive coding framework provides a significant amount of explanatory power to different areas of cortical function, such as the auditory system (Smith & Lewicki, 2006; Vuust et al., 2009), ventral midbrain (O'Doherty et al., 2006), hippocampus (Mehta, 2001) and the motor cortex (K. J. Friston et al., 2010). In summary, there lies a notable amount of evidence that supports the notion of the brian operating under a unifying computational strategy that follows the principles of predictive coding, namely that "[...] the brain actively predicts the hidden causes of incoming sensory information." (Huang & Rao, 2011, p. 591).

1.1.2 Experimental investigation of predictive coding

A variety of experimental setups have been employed in order to investigate how predictive coding manifests in various areas, using both neurobiological, brain imaging and behavioral measures (Keller et al., 2012; Modirshanechi et al., 2019; O'Doherty et al., 2006; Srinivasan et al., 1982; Visalli et al., 2021). A common denominator for a multitude of studies has been to utilize a setup that adheres to the oddball paradigm (e.g. Garrido et al., 2008; Huettel & McCarthy, 2004; Kim & McAuley, 2013; Kirihara et al., 2020; Moran et al., 2013; Schindel et al., 2011).

The reason for this can be explained by the inherent structure of the oddball paradigm, namely that it contains two types of stimuli; standard, which is the stimulus that occurs substantially more frequently than the other type of stimuli, and the deviant (also referred to as the oddball) (see Figure 4).
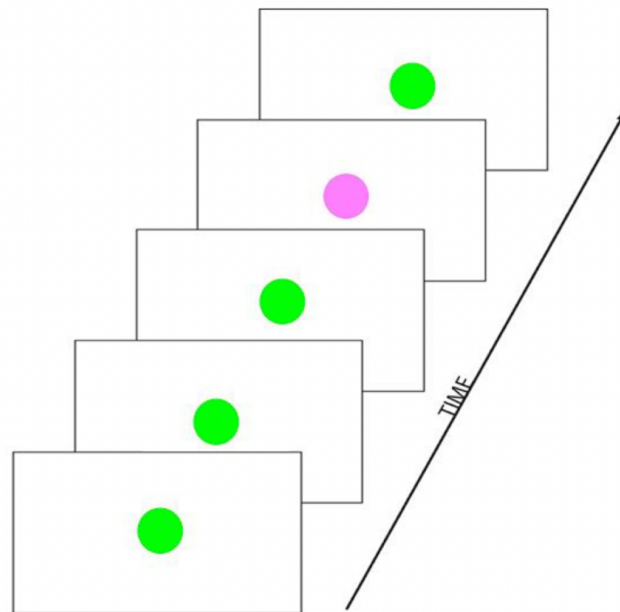
Figure 4: *An example of a visual oddball paradigm task. The stimuli are shown in temporal succession. In this example, the green ball is the standard stimulus while the pink is the deviant (Figure reprinted from Daltrozzo & Conway, 2014).*

Oddball paradigm tasks offer an ideal way to assess predictive coding in the brain since its structure facilitates the investigation of one of the main principles of the theory; that the brain constantly forms predictions of the next stimulus based on the previous stimuli (Kirihara et al., 2020; Modirshanechi et al., 2019). The paradigm thus enables researchers to analyze how the length of the sequence of standards preceding the deviant affects the response, which in the literature often is referred to as the prediction-error or the 'surprise' (Donchin, 1981; Garrido et al., 2008; Modirshanechi et al., 2019). The magnitude of the elicited surprise has been found to be positively correlated with the length of the preceding sequence of standards, supporting the hypothesis that the brain predicts the next stimulus based on the antecedent stimuli (Kirihara et al., 2020). One common method to measure the elicited prediction-error response is by applying the non-invasive, electrophysiological technique that is electroencephalography (EEG) (Garrido et al., 2008; Mousavi et al., 2020; Visalli et al., 2021).

EEG measures the electrical activity in microvolts produced by neuronal activity in the brain via electrodes positioned on top of the scalp and thus allows researchers to discern how the brain responds to a variety of stimuli. One of the advantageous attributes of EEG, compared to other brain imaging techniques such as

fMRI, is its high temporal resolution. It works within a timeframe of milliseconds and thus capacitates researchers to measure the brain's response to a given stimulus with a high temporal precision. One of the downsides to EEG lies within the fact that the activity is measured from the scalp, which results in poor spatial resolution. To circumvent this issue, one can time the onset of the stimulus presented to the subject with the EEG signal over numerous trials and subsequently perform signal averaging of all the electrode channels and trials. Through aligning the onset of the stimulus, the noise can thus be filtered out and it is possible to discern the electrical activity of the response related to the event (i.e. stimulus), which is referred to as the event-related potential (ERP) (Bear et al., 2020). ERPs are the main target of interest when studying predictive coding in the brain via EEG, since it is possible to see how the magnitude of some specific ERPs differ depending on whether the stimulus is standard or deviant. One of these ERP components is the positive late component, commonly referred to as P300, whose magnitude is used to gauge the size of the prediction-error or the subjective surprise elicited by the deviant in an oddball paradigm task (Donchin, 1981; Modirshanechi et al., 2019; Mousavi et al., 2020; Polich, 2007).

The definitions of the amplitude of the P300 varies across literature; a common choice is the mean of the EEG signal in a time interval centered around the positive peak located around 300 ms post-stimulus onset, also known as the windowed mean approach (Mousavi et al., 2020; Patel & Azzam, 2005). Evidence has suggested that at least a portion of the generation emanates from frontal and temporal/parietal activations (Ebmeier et al., 1995; Kirino et al., 2000) and is "[...] related to context-updating operations and subsequent memory storage" (Polich, 2007, p. 2136). More specifically, the neurobiological basis for the P300 generation has shown to be modality specific, where "[...] contributions come from the inferior temporal and superior parietal cortex for the visual and from the superior temporal cortex for the auditory modality." (Linden, 2005). Although more research is required for a comprehensive assessment of the neural basis of the P300 component.

1.2 Present study

In direct continuation of the theoretical foundations presented above, this pilot study will employ EEG to analyze the ERP component P300 in a visual oddball paradigm task. We will do so as to assess the actuality of the core premise of predictive coding, namely that the brain constructs a prediction of the ensuing stimulus actuated by the precursory stimuli. To enable this inquiry, the stimuli of the experimental setup will, in addition to the standard stimulus, consist of two types of deviants: one that deviates from the standard early on and another which deviates later. This will be actualized by having a stimulus that consists of a sequence of five digits. Based on principle coding and evidence supporting the framework, we predict that the magnitude of the P300 component will be greater for the deviant that deviates later in the sequence compared to the early deviant. We hold this supposition since the later deviant is preceded by more congruent digits than the early deviant, which should form a stronger prediction that the later deviant will be a standard trial – when the prediction is confuted, it should thus elicit a stronger prediction-error signal gauged by the magnitude of the amplitude of P300.

We will formalize the above argumentation into the following hypothesis:

H_2: *The magnitude of the amplitude of the P300 component will be greater for deviants that deviate later from the standard compared to earlier*

It should be noted that the hypothesis above builds on the assumption that there is a significant difference between the magnitude of the P300 for standard and deviant trials. We will therefore formulate a preliminary hypothesis, that is stated as the following:

H_1: *The magnitude of the amplitude of the P300 component will be greater for deviants compared to standards*

2. Methods

2.1 Experimental design

The experimental task was a variation of the oddball paradigm and comprised a learning phase and a recognition phase. In the introduction of the experiment, prior to the learning phase, the participant was informed that they would have to memorize a sequence of five digits that they had to imagine in their mind after the sequence was shown. This was repeated 30 times. After a 20-second break, the participant entered the recognition phase, where they had to either press '1' on the keyboard if the trial was congruent with the sequence learned in the memorization phase (i.e. the standard stimuli) or '2', if the stimulus was incongruent with the memorized sequence. We choose to have an active experiment (i.e. the participant had to respond to the stimuli, even though this behavioral response was not used in the analysis) as opposed to a passive experiment (i.e. the participant just had to passively watch the stimuli). This was chosen on the basis that active experiments have shown to be more successful with maintaining the subject's attention on the task (Mousavi et al., 2020). The recognition phase consisted of 90 trials and consisted of 3 types of stimuli: the standard sequence, a deviant that deviated from the standard at the second digit and a deviant that deviated at the fourth digit (see Figure 5). There were 30 of each type of stimuli, which was shown in a random order. For all sequences shown, each digit was shown for 0.35 seconds in continuous succession. The trial number was shown preparatory before and there was a 2-second pause with a blank screen after each trial.
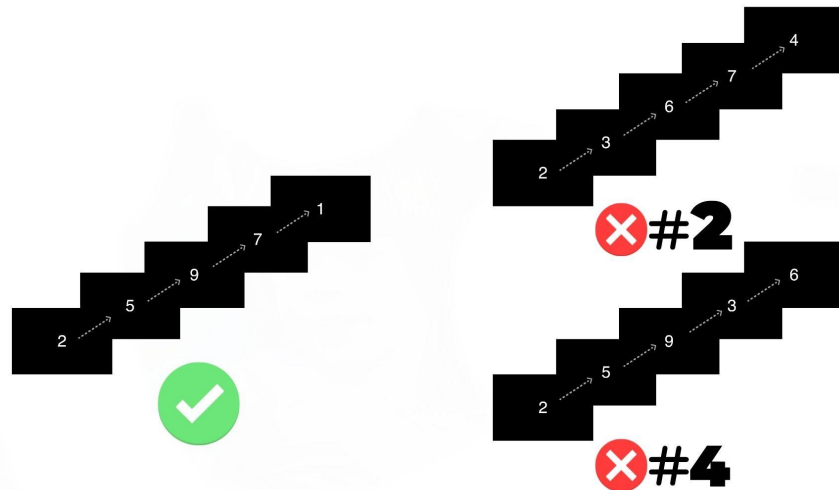
Figure 5: *The three different types of stimuli; the standard sequence, the deviant sequence where the second digit is incongruent and the deviant sequence where the fourth digit is incongruent.*

2.2 Procedure and apparatus

Due to limited resources and in concession to the present paper being a pilot study, only one participant participated in the experiment. The subject was a neurologically normal 23-year-old male that was recruited via the course 'Cognitive Neuroscience', which is a course placed during the 4th semester of the BSc programme of Cognitive Science at Aarhus University. The experiment was conducted at the EEG-research facility at Aarhus Universitetshospital, Skejby, where the subject, prior to the experiment, gave their informed consent to their participation. The participant then had a 32-electrode EEG cap mounted to their scalp in accordance with the international 10/20 system. Electrode 28 and 32 were positioned just under to the left of and over the left eye as to record the electrooculogram (EOG). Conductive gel was inserted into the mount holes of the electrodes before insertion to ensure good connectivity. Some electrodes, specifically the lateralized, indicated mediocre connectivity, but it was deemed to be acceptable for the time being. The experiment was carried out in an acoustical dampened room, as to ensure no disrupting sounds would interfere with the experiment and the distance from the participant's eyes to the screen was approximately 70 cm. The sampling rate of the EEG system, a BrainAmp amplifier system (Brain Products GmbH, Gilching, Germany), was set to 1000 Hz with the FCz position used as reference. Impedances of the electrodes were retained below 25 kΩ.

2.3 Preprocessing

The raw data was loaded into and preprocessed using the programming language Python 3 (Van Rossum & Drake, 2009) and the open-source Python package MNE (Gramfort et al., 2013). The signal from all the EEG-channels in the raw data was visually inspected, as to assess whether there were any 'bad' channels (i.e. channels that exhibited unusual amounts of noise) – no channels indicated any abnormalities. The raw data then went through a high-pass filtering specified to 0.1 Hz followed by a low-pass filter set to 40 Hz. This follows the common practices of EEG-preprocessing, as the high-pass filter minimizes slow drifts in the data (e.g. potentials produced by the skin of the scalp) while the low-pass filter excludes high-frequency noise (e.g. noise produced by lines in the experimental environment). Subsequently, an independent component analysis (ICA) was performed on the filtered data in order to detect and remove artifacts (e.g. eye-blinks) from the rest of the signal. The ICA revealed a component that seemed to capture the noise generated by the participant's eye blinks, assessed by the corresponding graph and topography. This component was then removed from a copy of the original raw data, whereafter the copy underwent the same filtering process as described above, thus substituting the former processed data. For the epoch generation, a rejection criteria was specified to exclude everything that had a maximum peak-to-peak signal amplitude (i.e. the absolute difference between the lowest and highest signal value for each channel) above 150 microvolts, since signals above this boundary are not thought to originate from cortical activity. The time interval for the epochs was tentatively set to a time interval from 200 ms prior to 500 ms after stimulus onset. In the final steps, the epochs were downsampled to 250 Hz, grouped into groups apposite to the analysis and afterwards averaged.
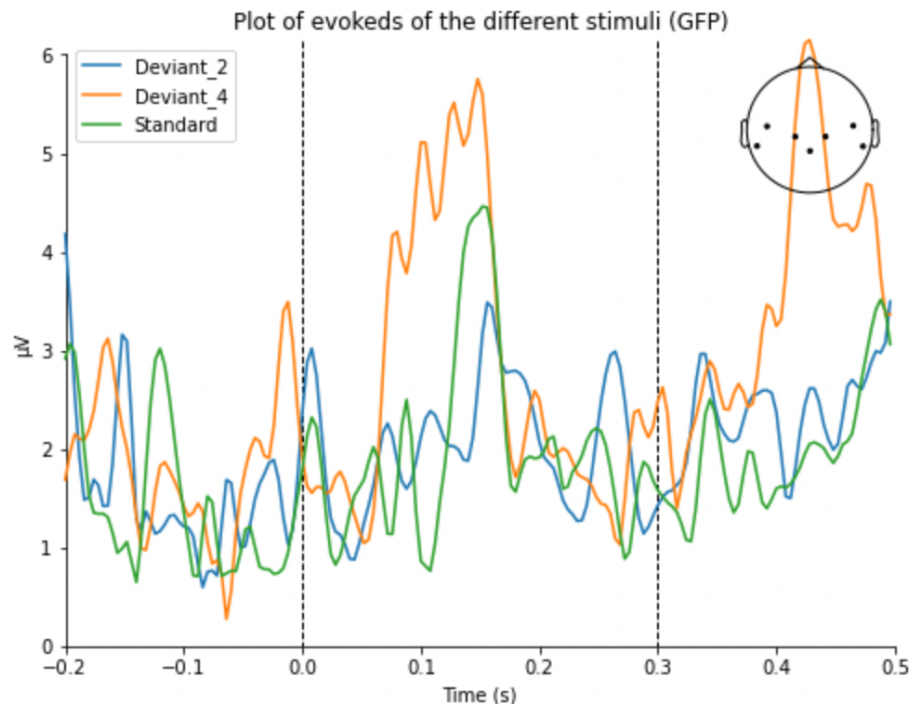
Figure 6: *Plot of the evokeds of the different stimuli (i.e. deviant deviating at digit 2, deviant deviating at digit 4 and standard). At 300 ms after stimulus onset, there appears to be a greater response for the deviant 4 compared to the other stimuli, which adheres to our hypothesis. The deviant 2 although crosses with the standard at 300 ms, with the standard eliciting a stronger response slightly before. At first glance, it therefore appears that the data does not fit with our hypothesis, but this will be further assessed in the analysis.*

2.4 Analysis

The chosen analytical approach employs the windowed mean of the evokeds of the different stimuli in order to test whether there is a statistically significant difference between the different stimuli within the time interval of interest. The latency interval was *a priori* set to be from 250 ms to 350 after stimulus onset, chosen on the theoretical grounds that the P300 component is not always elicited at exactly 300 ms but in proximity to this time point (Mousavi et al., 2020; Patel & Azzam, 2005). The EEG-channels included in the analysis, as to provide a more focal measurement, was chosen based on the literature on the neurological basis of the P300 generation in a visual modality, which indicates the areas of interest to be the inferior temporal and superior parietal cortex. On the basis of this, the following channels were chosen *a priori*: T7, T8, TP9, TP10, Pz, CP1, CP2 (following the international 10–20 system, see Sazgar & Young, 2019).

For the deviant stimuli, the evokeds for the first deviating digits were used. As for the standard, the evokeds of the second and fourth digit in the sequence were chosen, as to match the structure of the deviant stimuli since no literature to guide the choice could be found. To test the hypotheses, a t-test with an alpha-value set to 0.05 was first performed on the congregated deviants and the standard, followed by a t-test on the two types of deviants.

3.   Results

3.1 H_1

The result from the t-test performed on the deviants and the standard showed no statistical difference between the windowed mean amplitudes of the corresponding evokeds, $t = 0.391$, $p = 0.702$

3.2 H_2

The result from the t-test performed on the deviant with the second digit deviating and the deviant with the fourth digit deviating did not show any statistical difference between the windowed mean amplitudes of the corresponding evokeds, $t = 0.790$, $p = 0.445$.

4.   Discussion

4.1 Summary of results

Foremost, a t-test was conducted in order to investigate whether there was a statistical significant difference between the amplitude of the P300 component of the deviant stimuli and the standard stimulus, as to test the primlimary hypothesis "*H_1: The magnitude of the amplitude of the P300 component will be greater for deviants compared to standards*." The result of the test indicated that there was no difference between the amplitudes of the P300 for the two types of stimuli, signifying that the deviants did not give rise to a greater prediction-error compared to the standards. We therefore reject the hypothesis, thus diverging with otherwise robust evidence held within the literature of predictive coding. On the grounds of the foregoing, any potential significant results corroborating the other hypothesis will be discredited, since it builds upon the assumption that there indeed is a significant difference

between the deviants and standard in terms of elicited P300 magnitude. For punctilious reasons, we will continue with reviewing the results of the second hypothesis testing

An additional t-test was performed to evaluate the veracity of the principal hypothesis: "***H₂:*** *The magnitude of the amplitude of the P300 component will be greater for deviants that deviate later from the standard compared to earlier*". The statistical test produced a *p*-value that was closer to the set alpha-value than the previous, however still greatly insignificant. This entails that we reject the hypothesis, signifying that our results did not conform to our expectation that a sequence that deviates later would elicit a stronger prediction-error rather than if it deviated earlier.

In summary, none of our results provided evidence that substantiates predictive coding in the brain.

4.1 Limitations and further development

To delve into the reason why we did not obtain any results that support the predictive coding framework, not even for the preliminary hypothesis, numerous possibilities need to be considered. First and foremost, one apparent reason lies within the fact that the present study is overwhelmingly underpowered, since data was only obtained from one participant. Aware of this inherent shortcoming with pilot studies, the amount of trials in the recognition phase was maximized to get as much data as possible in an effort to counteract this flaw, evidently with no success. Since previous pilot studies have been able to have significant or close-to-significant results despite them being underpowered, other accounts need to be considered, as to refine the experimental setup for future studies.

Several potential drawbacks can be identified in the experimental procedure: firstly, multiple of the electrodes indicated insufficient connectivity, which is the basis for bad readings of the electrical impulses. In an effort to rectify these, additional conductive gel was added into the mounts of the bad electrodes, which seemed to solve the issue for an acceptable amount. Due to inexperience with setting up the measuring equipment, we were unaware of the caveat of adding a surplus of gel can result in bridges between the electrodes, which form the grounds of erroneous measuring, since the affected electrodes are no longer independent from each other. Another potential drawback lies within the circumstances of the

experimental run-through, namely that the participant had to wait for an extensive amount of time before the experiment could commence due to technical difficulties. This might have given rise to a decrement in the subject's attention, a concern that was confirmed by the subject themselves at the termination of the experiment. Since attention has been found to affect the magnitude of the P300 component elicited by the deviant (Polich, 2007), this might have had a confounding effect on the strength of the prediction-error response.

Finally, we need to consider whether the inherent structure of the experiment is the principal cause of the rejection of our hypotheses. Looking into the frequency of the different stimuli, our study deviates from other oddball paradigm tasks in terms of that the standard sequence occurred just as often as the deviants. Conventionally, the standard stimulus occurs predominantly more often that the deviant stimulus, which was not the case for our studies – in actuality, there were twice as many deviant trials than standard in the recognition phase, which can have caused the deviants to be just as, if not more expected than the standard stimulus. This has likely had substantial effects on the elicited prediction-error of the deviant stimuli. Additionally, our experimental setup and analysis focused on within-trial deviants as opposed to the between-trial surprise (i.e. we looked into the elicited prediction-error for whenever the sequence deviated from the standard, whereas conventional oddball tasks look into the prediction-error when the stimulus/trials deviates from the foregoing ones). To investigate how preceding observations affect the prediction of the upcoming stimulus, it might have been more efficacious to employ a roving oddball paradigm, where all types of stimuli can either be standard or deviant stimulus. The same stimulus is then repeated for varying lengths of sequences before switching to the other type of stimulus. This variation of the paradigm could therefore offer a more rigorous investigation of the present study's investigation of predictive coding.

5.   Conclusion

In this pilot study, we sought out to investigate predictive coding in the brain in terms of how it predicts the next stimulus based on the previous observations. The investigation was carried out by conducting an oddball paradigm experiment with three types of stimuli, that all consisted of a sequence of five digits: the standard, which was the 'correct' sequence, a deviant where the second digit deviated from the standard and a deviant where the fourth digit deviated from the standard. Based on the theory of predictive coding, it was hypothesized that the deviants would elicit a stronger prediction-error compared to the standard and that the deviant with the fourth digit deviating would produce a greater prediction-error than the deviant with the second digit deviating. As a proxy for the elicited prediction-error, the magnitude of the P300 component of the evokeds of the different stimuli was used in the statistical analysis to test the hypotheses. The analysis resulted in insignificant results, which led to the rejection of the hypotheses. Subsequently, different causes to the rejection of the hypotheses were considered and suggestions to improvements of the study were put forward.

6.   References

Bear, M., Connors, B., & Paradiso, M. A. (2020). *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain, Enhanced Edition*. Jones & Bartlett Learning.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Daltrozzo, J., & Conway, C. (2014). Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Frontiers in Human Neuroscience*, *8*, 437. https://doi.org/10.3389/fnhum.2014.00437

Donchin, E. (1981). Surprise!... Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Dong, D. W., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, *6*(3), 345–358. https://doi.org/10.1088/0954-898X_6_3_003

Ebmeier, K. P., Steele, J. D., MacKenzie, D. M., O'Carroll, R. E., Kydd, R. R., Glabus, M. F., Blackwood, D. H. R., Rugg, M. D., & Goodwin, G. M. (1995). Cognitive brain potentials and regional cerebral blood flow equivalents during two- and three-sound auditory "oddball tasks." *Electroencephalography and Clinical Neurophysiology*, *95*(6), 434–443. https://doi.org/10.1016/0013-4694(95)00173-5

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, *4*(12), 2379–2394. https://doi.org/10.1364/JOSAA.4.002379

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage*, *42*(2), 936–944. https://doi.org/10.1016/j.neuroimage.2008.05.018

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. S. (2013). MEG and EEG Data Analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(267), 1–13. https://doi.org/10.3389/fnins.2013.00267

Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, *436*(7047), 71–77. https://doi.org/10.1038/nature03689

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *WIREs Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Huettel, S. A., & McCarthy, G. (2004). What is odd in the oddball task?: Prefrontal cortex is activated by dynamic changes in response strategy.

Neuropsychologia, *42*(3), 379–386.

https://doi.org/10.1016/j.neuropsychologia.2003.07.009

Keller, G. B., Bonhoeffer, T., & Hübener, M. (2012). Sensorimotor Mismatch Signals

in Primary Visual Cortex of the Behaving Mouse. *Neuron*, *74*(5), 809–815.

https://doi.org/10.1016/j.neuron.2012.03.040

Kim, E., & McAuley, J. D. (2013). Effects of pitch distance and likelihood on the

perceived duration of deviant auditory events. *Attention, Perception, &

Psychophysics*, *75*(7), 1547–1558.

https://doi.org/10.3758/s13414-013-0490-5

Kirihara, K., Tada, M., Koshiyama, D., Fujioka, M., Usui, K., Araki, T., & Kasai, K.

(2020). A Predictive Coding Perspective on Mismatch Negativity Impairment

in Schizophrenia. *Frontiers in Psychiatry*, *11*, 660.

https://doi.org/10.3389/fpsyt.2020.00660

Kirino, E., Belger, A., Goldman-Rakic, P., & McCarthy, G. (2000). Prefrontal

Activation Evoked by Infrequent Target and Novel Stimuli in a Visual Target

Detection Task: An Event-Related Functional Magnetic Resonance Imaging

Study. *Journal of Neuroscience*, *20*(17), 6612–6618.

https://doi.org/10.1523/JNEUROSCI.20-17-06612.2000

Linden, D. E. J. (2005). The P300: Where in the Brain Is It Produced and What Does

It Tell Us? *The Neuroscientist*, *11*(6), 563–576.

https://doi.org/10.1177/1073858405280524

Mehta, M. R. (2001). Neuronal Dynamics of Predictive Coding. *The Neuroscientist*,

7(6), 490–495. https://doi.org/10.1177/107385840100700605

Millidge, B., Seth, A., & Buckley, C. L. (2022). *Predictive Coding: A Theoretical and

Experimental Review* (arXiv:2107.12979). arXiv.

http://arxiv.org/abs/2107.12979

Modirshanechi, A., Kiani, M. M., & Aghajan, H. (2019). Trial-by-trial

surprise-decoding model for visual and auditory binary oddball tasks.

NeuroImage, *196*, 302–317.

https://doi.org/10.1016/j.neuroimage.2019.04.028

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., & Friston, K. J.

(2013). Free Energy, Precision and Learning: The Role of Cholinergic

Neuromodulation. *The Journal of Neuroscience*, *33*(19), 8227–8236.

https://doi.org/10.1523/JNEUROSCI.4255-12.2013

Mousavi, Z., Entesari, T., & Aghajan, H. (2020). Prediction of the Brain Surprise in

Sequences of Oddball Stimuli. *2020 28th Iranian Conference on Electrical

Engineering (ICEE)*, 1–6. https://doi.org/10.1109/ICEE50131.2020.9260688

O'Doherty, J. P., Buchanan, T. W., Seymour, B., & Dolan, R. J. (2006). Predictive

Neural Coding of Reward Preference Involves Dissociable Responses in

Human Ventral Midbrain and Ventral Striatum. *Neuron*, *49*(1), 157–166.

https://doi.org/10.1016/j.neuron.2005.11.014

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected

Studies of the Event-Related Potential. *International Journal of Medical

Sciences*, *2*(4), 147–154.

Patton, L. (2018). Hermann von Helmholtz. In E. N. Zalta (Ed.), *The Stanford

Encyclopedia of Philosophy* (Winter 2018). Metaphysics Research Lab,

Stanford University.

https://plato.stanford.edu/archives/win2018/entries/hermann-helmholtz/

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical

Neurophysiology*, *118*(10), 2128–2148.

https://doi.org/10.1016/j.clinph.2007.04.019

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

Sazgar, M., & Young, M. G. (2019). Overview of EEG, Electrode Placement, and Montages. In M. Sazgar & M. G. Young (Eds.), *Absolute Epilepsy and EEG Rotation Review: Essentials for Trainees* (pp. 117–125). Springer International Publishing. https://doi.org/10.1007/978-3-030-03511-2_5

Schindel, R., Rowlands, J., & Arnold, D. H. (2011). The oddball effect: Perceived duration and predictive coding. *Journal of Vision, 11*(2), 17. https://doi.org/10.1167/11.2.17

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature, 439*(7079), 978–982. https://doi.org/10.1038/nature04485

Spratling, M. W. (2010). Predictive Coding as a Model of Response Properties in Cortical Area V1. *Journal of Neuroscience, 30*(9), 3531–3543. https://doi.org/10.1523/JNEUROSCI.4911-09.2010

Srinivasan, M. V., Laughlin, S. B., Dubs, A., & Horridge, G. 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

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Visalli, A., Capizzi, M., Ambrosini, E., Kopp, B., & Vallesi, A. (2021). Electroencephalographic correlates of temporal Bayesian belief updating and surprise. *NeuroImage, 231*, 117867. https://doi.org/10.1016/j.neuroimage.2021.117867

Otto Sejrskild Santesson (202004347)

202004347@post.au.dk

31st of May, 2022

BSc Cognitive Science, Aarhus University

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009).

Predictive coding of music – Brain responses to rhythmic incongruity. *Cortex*,

45(1), 80–92. https://doi.org/10.1016/j.cortex.2008.05.014