

**What happens next versus when “next” happens:
Dissociating spatial and temporal prediction mechanisms**

by

Dean R. Wyatte

B.S., Indiana University Bloomington, 2007

M.A., University of Colorado Boulder, 2010

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Psychology and Neuroscience

2014

This thesis entitled:
What happens next versus when “next” happens:
Dissociating spatial and temporal prediction mechanisms
written by Dean R. Wyatte
has been approved for the Department of Psychology and Neuroscience

Randall C. O'Reilly

Prof. Tim Curran

Prof. Albert Kim

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Wyatte, Dean R. (Ph.D., Cognitive Neuroscience)

What happens next versus when “next” happens:

Dissociating spatial and temporal prediction mechanisms

Thesis directed by Prof. Randall C. O'Reilly

Contents

Tables

Figures

Chapter 1

1.1 Introduction

How does the brain predict what happens from one moment to the next? This is an important question in research on perception across all modalities that is surprisingly often overlooked in the fields of psychology and neuroscience. For example, most experiments are designed to measure evoked responses to a randomly chosen, isolated stimulus under the tacit assumption that response variability is irrelevant noise that should be averaged out across many presentations. Computational models of perceptual processing often operate under similar assumptions in which stimuli are presented as random “snapshots” from which some common set of features should be learned to minimize representational variability across presentations (e.g., SerreOliva-Poggio07MutchLowe08; although see Foldiak91, for a notable exception). These experimental and computational assumptions stand in contrast to the event structure of the physical world, which is deterministic from one moment to the next.

An equally important question is concerned with how the brain knows when to make its next prediction. Predicting what happens next requires integrating information over some time frame and using the result to drive the actual prediction, but when should integration start? And for how long? Perceptual processing has been shown to undergo temporal fluctuations and in extreme cases, stimuli can be rendered imperceptible if presented during one of these fluctuations (BuschDuboisVanRullen09VanRullenBuschDrewesEtAl11; MathewsonLlerasBeckEtAl11). Again, many laboratory experiments tend not to be concerned with these temporal fluctuations, as

they simply add variability to responses that will average out over a large number of trials and can be mitigated by design decisions such as using a fixation cross to denote the start of a trial. Computational models, similarly, often completely ignore time altogether, although recent advances in spiking models of perceptual processing (e.g., MasquelierThorpe07) are beginning to address this issue.

The goal of the proposed work is to begin systematically investigating the neural mechanisms and circuitry related to prediction and temporal integration. Recently, our lab has developed a theoretical framework and general model for describing how these functions are implemented in the brain. This framework, henceforth referred to as *LeabraTI*,¹ brings together a large number of independent findings from the systems neuroscience literature to describe how multiple interacting mechanisms accomplish prediction and temporal integration in cortex. The fundamental proposal of the LeabraTI theory is that time in the brain is discretized (at least partially) into reference frames that can be associated via general learning mechanisms (e.g., Hebbian and error-driven learning) so that representation of information during one frame can be used to predict what happens during the next. This proposal requires at least two mechanisms: (1.) A mechanism that establishes reference frames over which information is integrated and, (2.) A mechanism that generates the actual prediction itself and validates it against what actually happens.

As will be discussed in detail in Chapter ??, the LeabraTI theory's proposed mechanisms are suggested to be dissociable, generating signatures at distinct spectral frequencies which can be measured physiologically. The potential dissociability of the spatial and temporal components of prediction establishes a number of immediately testable predictions that will form the experimental component of the proposed work.

¹ Leabra refers to a general model of learning in the neocortex (O'ReillyMunakata00O'ReillyMunakataFrankEtAl12); TI to Temporal Integration.

Chapter 2

2.1 Introduction

The *LeabraTI* framework is a mechanistic description and general model of how prediction and temporal integration might work in the brain. The general idea is closely related to the Simple Recurrent Network (SRN) (Elman90Servan-SchreiberCleeremansMcClelland91) an artificial neural network architecture that explicitly represents temporally lagged information in discrete “context” units whose activity gets integrated with more current information to predict what happens in the next time step (Figure ??a). This method of copying a contextual representation from an intermediate representation at discrete intervals was originally shown to be a robust way to leverage error-driven learning to represent latent temporal structure in auditory streams and artificial grammars. More generally, the SRN’s explicit representation of temporal context can capture the latent structure of any stimulus that varies systematically over time, making it a good basis for a generic prediction and temporal integration mechanism.

However, *LeabraTI* differs in several key ways from the traditional SRN architecture, primarily in the way temporally lagged context is represented and used in neural processing. These differences are due to biological constraints imposed by the circuitry of the neocortex, and thus form a number of testable predictions that can be used to evaluate the validity of the *LeabraTI* framework. The central prediction of *LeabraTI* is that temporal context is updated in a periodic manner, approximately every 100 ms, subserved by deep (Layer 5 and 6) neurons. This periodic updating is suggested to contribute to the brain’s alpha rhythm, which has been studied extensively

using scalp electroencephalography. This chapter **TODO – also need something here about error driven learning**

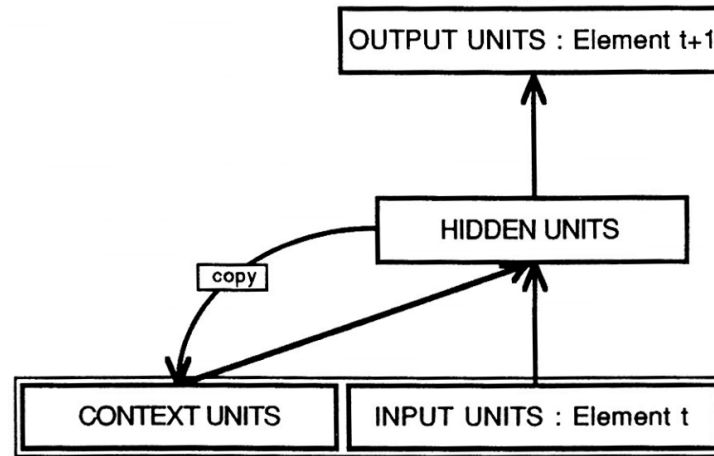
2.2 LeabraTI biological details

2.2.1 Laminar structure and microcircuitry of the neocortex

A salient feature of the brain, and potential clue in realizing how an SRN-like computation might be carried out in biological neural circuits, is the laminar structure prevalent across the neocortex (Figure ??b). Incoming information from the sensory periphery is transmitted through the thalamus and targets Layer 4 neurons in the primary sensory cortices (e.g., V1). From there, Layer 4 neurons propagate spikes to superficial neurons (Layers 2 and 3) which in turn target Layer 4 neurons of higher-level cortices, forming the prominent corticocortical feedforward pathways that subserve visual and auditory recognition (FellemanVanEssen91). Corticocortical feedback originates in superficial layers or Layer 6 of the higher-level cortex and generally terminates on superficial neurons of the lower-level cortex (RocklandPandya79). In addition to these interareal pathways, there exists a canonical microcircuit of the form Layer 4 \rightarrow Layer 2/3 \rightarrow Layer 5 \rightarrow Layer 6 that routes spike propagation through the local neuronal structure (DouglasMartin04ThomsonLamy07). This microcircuit forms the core computational unit of LeabraTI, as will be described in this and the following sections.

The importance of the local microcircuit was first suggested by Vernon Mountcastle in his proposal regarding the gross columnar organization of the neocortex (see Mountcastle97, for a comprehensive review). Mountcastle's proposal states that microcolumns composed of around 80-100 neurons extending vertically through all six lamina with canonical circuitry form the core repeating structure of the neocortex. Neurons within a single microcolumnar circuit possess nearly identical receptive field tunings across lamina while neurons in neighboring microcolumns (radial separation greater than 600 μm) possess very different receptive field tunings but contribute to the higher-order macrocolumn (i.e., hypercolumn) structure (HubelWiesel77Jones00). Microcolumns

a)



b)

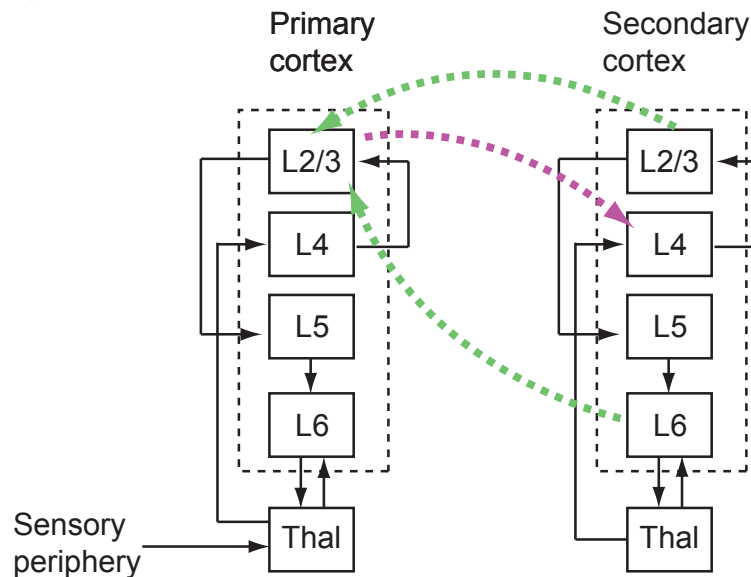


Figure 2.1: The Simple Recurrent Network (SRN) and microcircuitry of the neocortex. **a)** The SRN represents temporal information explicitly using discrete context units that are updated once per time step. Context is integrated with more current inputs to predict information at the subsequent time step. Reproduced from Servan-SchreiberCleeremansMcClelland91). **b)** The neocortex is laminated with canonical circuitry between neurons across layers and between areas. Intraareal connections are shown in black with interareal feedforward connections in purple and feedback connections in green.

have been identified in a variety of neural systems with this electrophysiological mapping and are also prominently visible under Nissl staining. Despite this evidence for their structural existence, any function of the microcolumn aside from an organizing principle remains debated (BuxhoevedenCasanova02HortonAdams05).

LeabraTI provides a computational role for the microcolumn, by mapping an SRN-like computation onto their Layer 4 \rightarrow Layer 2/3 \rightarrow Layer 5 \rightarrow Layer 6 circuit. In this mapping, superficial neurons continuously integrate feedforward and feedback interareal synapses to process current information. Layer 2/3 \rightarrow Layer 5 \rightarrow Layer 6 provides an intraareal pathway for explicitly representing temporal context deep layers, which are relatively isolated from nonlocal inputs. There is also appropriate circuitry for recirculating this context through the local microcolumn via Layer 4 to drive the learning of temporal associations. This basic idea provides a concise explanation for the strong degree of isotuning throughout a single microcolumn, as deep neurons need to represent the same information as superficial neurons except at a delayed interval.

More generally, LeabraTI's dichotomy of continuous integration in superficial layers and periodic updating of deep layers receives strong support by the literature. Recent studies that have employed depth electrodes to simultaneously record from multiple layers within a patch of cortex have indicated that superficial layers exhibit spectral power at much higher frequencies than deep layers. BuffaloFriesLandmanEtAl11) recorded responses from ventral visual sites V1, V2, and V4 in awake, behaving monkeys during a simple directed attention task, finding a dissociation in spike coherence frequency in superficial (gamma spectrum, peak ~ 50 Hz) and deep layers (alpha spectrum, peak ~ 10 Hz). A similar experimental paradigm expands on these findings by demonstrating cross-frequency coupling between gamma and alpha spectra localized to superficial and deep layers, respectively (SpaakBonfondMaierEtAl12). The cross-frequency coupling is characterized by a clear nesting of gamma activity within alpha cycles, suggesting that deep neurons' alpha activity might subserve a general pacemaker mechanism. In the context of LeabraTI, this pacemaker property is important to ensure the regular updating of context through deep layers and temporally predictable reintegration with more current information.

In summary, the laminocolumnar organization of the neocortex provides the dual pathways necessary for continuous information processing and the SRN's explicit temporal context representation. One question that remains, however, concerns the 10 Hz alpha periodicity of deep neurons. The Layer 4 → Layer 2/3 → Layer 5 → Layer 6 microcircuit only contains four synapses including the thalamus and intracolumnar monosynaptic latencies for regular spiking neurons are on the order of 5 ms or faster (Armstrong-JamesFoxDas-Gupta92LumerEdelmanTononi97). This relatively small amount of tissue, if driven with constant input, would cause deep neurons to spike at a rate much faster than 10 Hz. How such a circuit could produce the strong alpha power observed in recent depth recordings is described next.

2.2.2 Pacemaker properties of Layer 5 and thalamic bursting neurons

Layer 5 neurons can be roughly divided into 5a and 5b subtypes (ThomsonLamy07). Layer 5a neurons have relatively small cell bodies and exhibit “regular spiking” depolarization responses. They collect input from other Layer 5a neurons and pass it to 5b neurons and thus, likely play a simple information integration role. Layer 5b neurons, in contrast, have larger cell bodies and exhibit “intrinsic bursting” properties at ~10 Hz when over threshold (ConnorsGutnickPrince82SilvaAmitaiConnors91; FranceschettiGuatteoPanzicaEtAl95).

Thalamic neurons...

Layer 5b neurons project to Layer 6 whose neurons can also be roughly divided into corticocortical (CC) and corticothalamic (CT) subtypes (Thomson10).

Both Layer Layer 6 CC neurons have properties similar to Layer 5a neurons – they collect inputs from other Layer 6a neurons and pass it to Layer 6 CT neurons. Layer 6 CT neurons project specifically to the thalamus and also receive direct thalamic input forming a small circuit. They have

This rhythmic firing has been shown to persist even with constant sensory stimulation *in vivo* (LuczakBarthoHarris13), suggesting that Layer 5 neurons' alpha rhythmicity could implement a roughly 10 Hz gating function for spikes relayed to Layer 6 neurons.

Thus, Layer 6 specifically becomes the neural substrate of the SRN's temporally lagged context representation, representing information that is, on average, one alpha cycle (approximately 100 ms) in the past. This contextual storage occurs at an automatic interval due to the intrinsic pacemaking properties of Layer 5 neurons, and might implement a reference frame that essentially would allow the brain to know *when* to anticipate inputs. As such, intrinsic oscillations have been shown to phase lock to environmental stimulation (WillBerg07LakatosKarmosMehtaEtAl08SchroederLakatos09S ensuring environmental events coincide with key events like Layer 5 bursts in cortex.

Layer 6 sends axons toward the thalamus, completing the microcircuit within the local column and allowing the temporally lagged Layer 6 information to integrate with more current Layer 4 inputs. There also exists a direct connection between Layer 6 and Layer 4, that could be used for this purpose, although it has been noted as being weak compared to other intracolumnar connections (HirschMartinez06b). In either case, temporal associations could be created by simple Hebbian learning mechanisms that track high probability co-occurrences across past and present events.

The Leabra algorithm (OReillyMunakata00OReillyMunakataFrankEtAl12), however, also makes use of powerful error-driven learning (in addition to more standard Hebbian learning). In the context of temporal integration, error-driven learning would allow computation of error signals based on the difference between what is predicted to happen at a given moment (given the previous moments context as an input) and what actually happens. However, this computation requires that both the prediction and the actual sensation are represented subsequently within a single alpha cycle, which is not possible if the sensory periphery is always transmitting incoming inputs. To resolve this issue, the LeabraTI framework posits a mechanism to modulate or even block the transmission of inputs from the sensory periphery. A subset of cells in the thalamus exhibit alpha spectrum bursting properties similar to that of Layer 5 neurons (LopesdaSilva91; HughesLorincz-CopeEtAl04; LorinczCrunelliHughes08; LorinczKekesiJuhaszEtAl09), and thus perhaps perform a similar gating function. Specifically, these neurons appear to shift the balance of inputs to Layer 4 and superficial neurons between exogenous environmental inputs and endogenous inputs local to

the microcolumn.

When environmental inputs are downmodulated or blocked, Layer 6 context relayed via the thalamus is the dominant input to the microcolumn, which can be used to predict the incoming sensory event during the latter part of the alpha cycle. Importantly, during both the prediction and sensation phases, feedforward and feedback projections are constantly transmitting between lower and higher cortical areas. As previously mentioned, these projections originate and terminate predominantly in superficial layers, boosting their spike coherence to higher frequency spectra. This could potentially explain the differentially high gamma power in superficial layers compared to deep layers, and provides a compelling link between gamma oscillations and predicting specific details about the next sensory event.

2.2.3 Summary of LeabraTI computation

The overall computation of LeabraTI is shown in Figure ?? and summarized here. When thalamic cells burst (roughly every 100 ms), information from the sensory periphery is the primary driving force for Layer 4 neurons in primary cortex. This information is relayed downstream to higher-level cortical areas via the strong feedforward Layer 4 → Layer 2/3 → Layer 4 pathway (FellemanVanEssen91). Within the local microcolumn, Layer 5 neurons integrate this information, until thalamic bursting quiets (generally around 50 ms). At this point, Layer 5 cells burst, sending outputs to Layer 6 and shifting and inputs to the microcolumn endogenously. The information represented by Layer 6 neurons is temporally lagged (from the previous 50 ms) and is relayed to Layer 4 via non-bursting (regular spiking) thalamic neurons or via the direct Layer 6 → Layer 4 connection (not pictured in Figure ??), and might be maintained by reciprocal thalamocortical drive back to Layer 6. This information can be used as a prediction as to what will happen next when thalamic bursting resumes and veridical sensory information serves as the input once again. In the context of Leabra's error-driven learning these two phases correspond to the plus phase (sensation) and minus phase (prediction), which can be used to compute a sensory prediction error signal. This error signal modifies Layer 5 → Layer 6 synapses to minimize differences between

Figure 2.2: The LeabraTI model computation.

2.3 LeabraTI predictions

2.3.1 Relation to existing models

2.3.2 Testable predictions

TODO

Chapter 3

3.1 Introduction

TODO

3.2 Methods

3.2.1 Participants

A total of 29 students from the University of Colorado Boulder participated in the experiment (ages 18-26 years, mean=20.13; 16 male, 13 female). All participants were right-handed and reported normal or corrected-to-normal vision. Participants either received course credit or payment of \$15 per hour as compensation for their participation. Informed consent was obtained from each participant prior to the experiment in accordance with Institutional Review Board policy at the University of Colorado.

3.2.2 Stimuli

Novel “paper clip” objects similar to those used in previous studies (BulthoffEdelman92EdelmanBulthoff92) were created using MATLAB. Eight vertices were placed randomly on the surface of a sphere of unit radius and then joined together with line segments. The last and first vertex were also joined to form a closed loop so that line segment terminations were not a salient feature (BalasSinha09b). Objects were constrained to exclude extremely acute angles between successive segments (less

than 20 degrees) and were approximately rotationally balanced (center of mass within 10% of the origin). Objects were rotated completely about their vertical axis in steps of 12 degrees and rendered to bitmap images under an orthographic projection. A total of 16 objects were created using this procedure, yielding 480 images (30 images per object). Object examples are shown in Figure ??.

3.2.3 Procedure

Participants observed an entraining sequence of rotated views of a random object and performed a same-different judgement about a probe stimulus. On each trial, a view was randomly selected as the first view of the sequence followed by seven additional views spaced 24 degrees apart (Figure ??A, blue tick marks). Thus, the eight view entraining sequence spanned 168 degrees of the object. The entraining sequence was either presented in order (i.e., spatially predictable) or randomized. Following the entraining sequence after a 200 ms blank was a probe stimulus consisting of either an unseen view from the entraining object or a novel distractor. Unseen views were randomly sampled from the 12 degree interpolations between views of the entraining sequence (Figure ??A, magenta tick marks) and from outside of the span of the entraining sequence in increments of 24 degrees (Figure ??A, green tick marks).

Distractors were created from the original target objects by randomly selecting new spherical coordinates for six of the eight vertices and re-rendering them to bitmap images using the same method as the original target objects (12 degree steps about the vertical axis). Distractors conformed to the same constraints as the original target objects (no extremely acute angles, approximately rotationally balanced). Participants were instructed to respond “same” if they believed the probe depicted the same object as the entraining sequence or “different” if it depicted a distractor object. Participants received feedback after each trial according to whether their response was correct or incorrect. Responses were collected via a millisecond-accurate response box connected through the display computer’s serial port.

During the entraining sequence, object views were presented for 50 ms at either 10 Hz (i.e.,

temporally predictable) or at a variable rate by manipulating the interstimulus interval (ISI) between subsequent views. Temporally predictable ISIs were 50 ms, totaling 350 ms across the entraining sequence. Variable ISIs were selected by randomly generating seven ISIs that also summed to 350 ms (Figure ??B). ISIs were in the range of 16.67 ms (minimum) to 216.67 ms (maximum) in increments of 16.67 ms. Temporal unpredictability was maximized by generating 400 such ISI sequences, calculating the summed squared error (SSE) across subsequent ISIs in a sequence, and selecting the 100 sequences with the highest SSE for use during the experiment.

The experiment was displayed on an LCD monitor with a native resolution of 1280x1024 operating at 60 Hz using the Psychophysics Toolbox Version 3 (Brainard97Pelli97). Stimuli were presented on an isoluminant gray background and subtended approximately 5 degrees of visual angle. Trials began with a fixation cross (200 ms) followed by a blank (400 ms), the entraining sequence (750 ms total), a second blank (200 ms), and ended with the probe stimulus (50 ms) (Figure ??C). Participants were required to respond within 2000 ms. Trials were separated by a variable intertrial interval of 2000-2400 ms. The experiment contained 500 trials with an additional 20 practice trials that contained a longer blank (1000 ms) between the entraining sequence and the probe to familiarize subjects with the order of events during trials. Participants completed the 20 practice trials (which were discarded from analysis) prior to performing the 500 experimental trials.

3.2.4 EEG recording and preprocessing

The EEG was recorded using a 128 channel net (HCGSN 130) amplified through 200 M Ω amplifiers (Net Amps 200) and was sampled at 250 Hz. Impedances for each electrode were adjusted to less than 40 k Ω before and during the recording.

EEG data were preprocessed using the FieldTrip toolbox (OostenveldFriesMarisEtAl11). Raw data were first band-pass filtered between 1 Hz and 100 Hz with a 59-61 Hz band-stop and then epoched into 2350 ms segments that spanned the start of the pre-trial blank to 1000 ms after the probe stimulus. Individual segments were visually inspected and rejected if found to contain

muscle artifacts or atypical noise. Bad channels were also identified and temporarily removed from the data before performing ICA decomposition (DelormeMakeig04) to remove of ocular artifacts. Components related to ocular artifacts were identified based on their topographical distribution across electrodes. The data were reconstructed without the ocular components and any bad channels were replaced using spherical spline interpolation (PerrinPernierBertrandEtAl89). The resulting segments were re-referenced to the average reference.

3.2.5 Event-related averaging

Event-related averaging was performed separately for the entraining sequence, the 200 ms blank between the entraining sequence and the probe, and the probe itself. For the entraining sequence, the data were baseline corrected

All event-related averaging was further averaged over pools of seven electrodes centered over locations from the 10-10 system that are commonly associated with perceptual processing (e.g., DohertyRaoMesulamEtAl05RohenkohlNobre11). These pools included one midline site (OZ) and three sites over each hemisphere (O1/O2, PO3/PO4, and PO7/PO8) (Figure ??).

3.2.6 Time-frequency analysis

3.3 Results

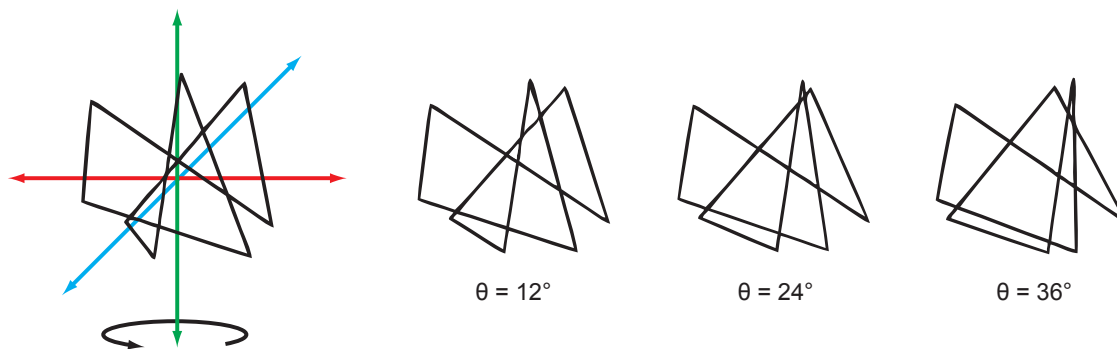
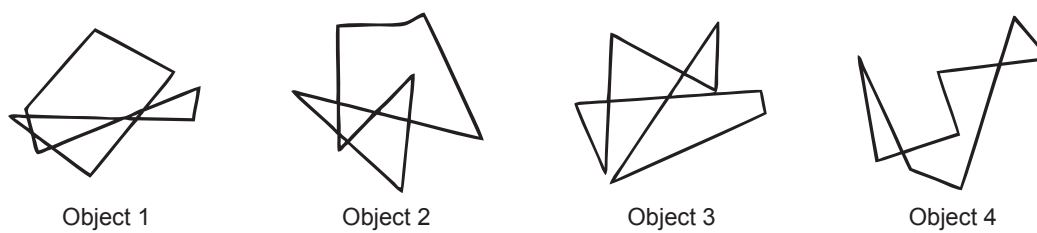
A**B**

Figure 3.1: Novel “paper clip” objects.

A: B:

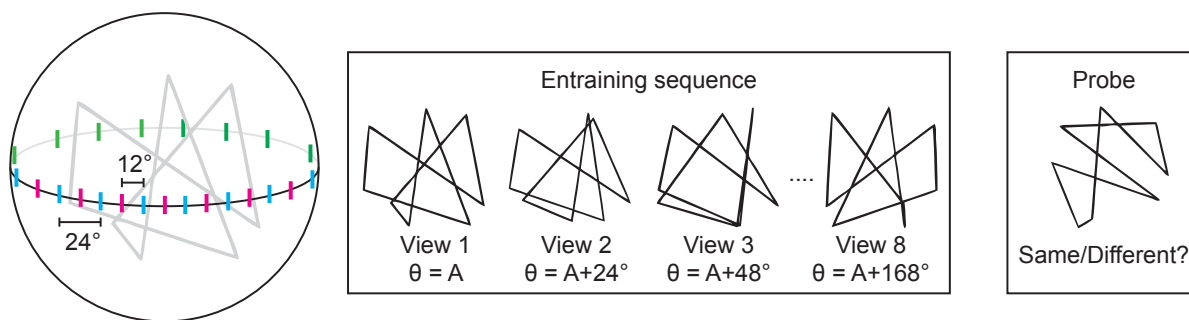
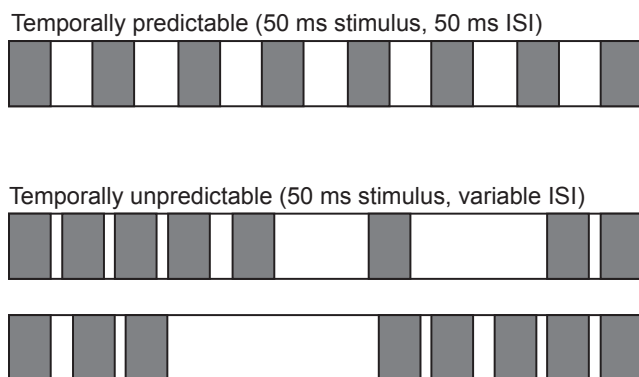
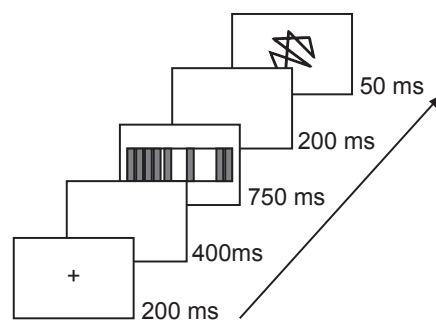
A**B****C**

Figure 3.2: Experimental procedure.

A: B: C:

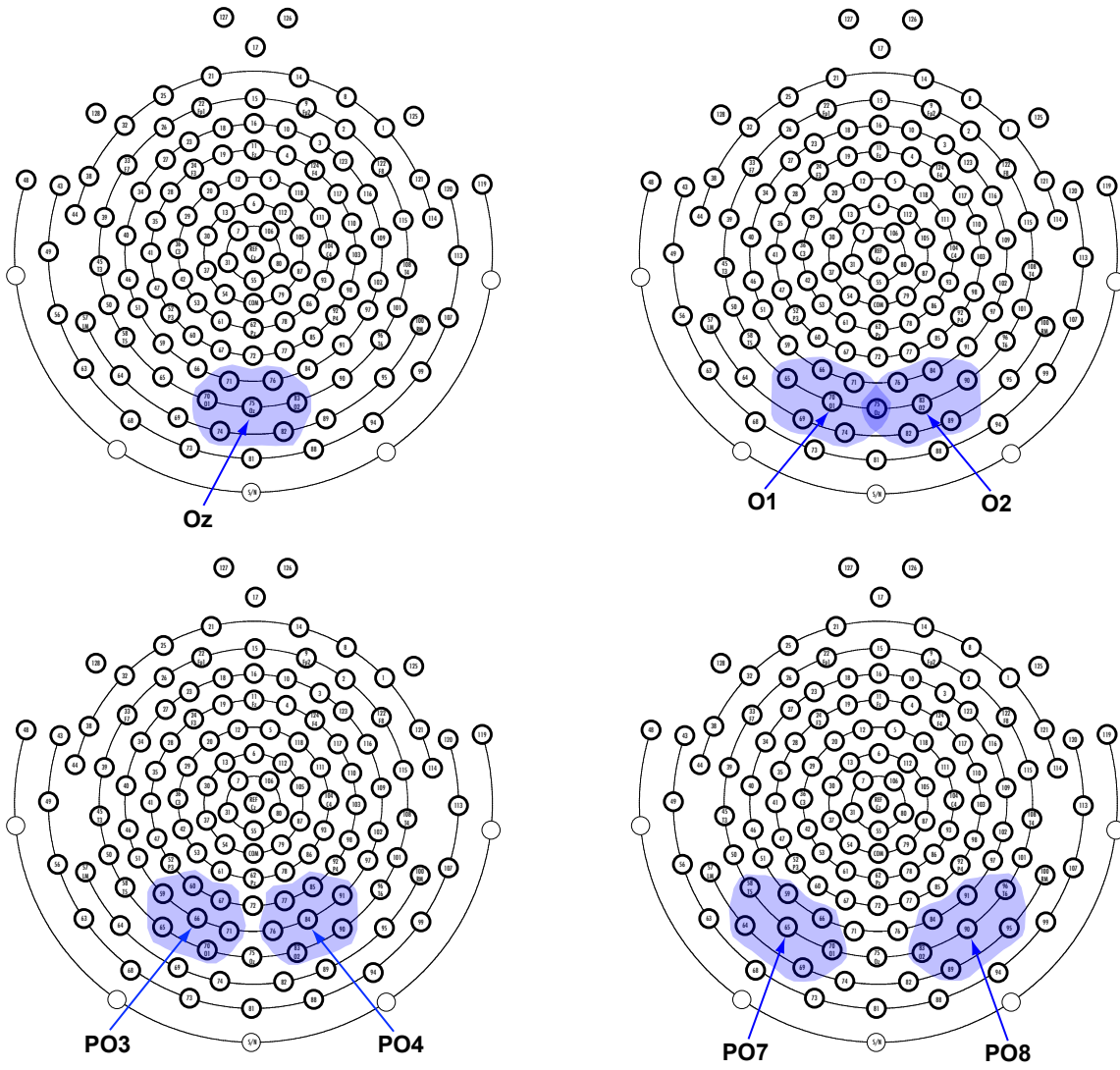


Figure 3.3: Electrode poolings.

Chapter 4

4.1 Introduction

TODO

4.2 Methods

4.2.1 Participants

A total of NN students from the University of Colorado Boulder participated in the experiment (ages NN-NN, mean=NN.NN years; NN male, NN female). All participants reported normal or corrected-to-normal vision and received course credit as compensation for their participation. Informed consent was obtained from each participant prior to the experiment in accordance with Institutional Review Board policy at the University of Colorado.

4.2.2 Stimuli

Novel “paper clip” objects (see REF PLEAST) were used as stimuli. A total of eight objects were created using the same method as in Experiment PLEAST. Four of the objects were target objects from Experiment PLEAST.

4.2.3 Procedure