What happens next and when "next" happens:
Mechanisms of spatial and temporal prediction

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 and when "next" happens:
Mechanisms of spatial and temporal prediction
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 and when “next” happens:

Mechanisms of spatial and temporal prediction

Thesis directed by Prof. Randall C. O’Reilly

Acknowledgements

Contents

Chapter

Figures

Figure

Chapter 1

Spatial and temporal prediction during novel object recognition: Temporal and spectral signatures

1.1 Introduction

How does the brain integrate information from one moment to the next and use it it to drive predictions about what will happen? A number of models of general neocortical function have highlighted the centrality of prediction in neural processing (e.g., Dayan, Hinton, Neal, & Zemel, 1995; Rao & Ballard, 1999; Lee & Mumford, 2003; Friston, 2005; George & Hawkins, 2009), but it is surprisingly but often overlooked in psychology and neuroscience investigations of sensory processing. Predictability might be useful in perceptual processes like object recognition, for example, by allowing better integration across features extracted over the course of several samples (Foldiak, 1991; Stringer, Perry, Rolls, & Proske, 2006; Wallis & Baddeley, 1997; Isik, Leibo, & Poggio, 2012). The timing of predictions might also be important. The LeabraTI model (Chapter **??**) as well as several other theories of sensory prediction (Arnal & Giraud, 2012; Giraud & Poeppel, 2012) emphasize that predictions occur in a pacemaker manner at regular intervals and thus temporal properties of endogenous processes as well as exogenous stimulation might also affect perceptual processing.

Previous work has implicated 7-13 Hz alpha oscillations in the allocation of spatial attention and anticipation of the temporal onset of stimuli (Gould, Rushworth, & Nobre, 2011; Belyusar et al., 2013; Rohenkohl & Nobre, 2011). Specifically, alpha oscillations desynchronize in the hemisphere contralateral to the attended region of visual space as well as in anticipation of the

onset of a time-varying stimulus. Pre-stimulus alpha oscillations have also been related to the detection rate of at-threshold point-light targets (Mathewson, Gratton, Fabiani, Beck, & Ro, 2009; Busch, Dubois, & VanRullen, 2009), suggesting that spontaneous fluctuations in alpha oscillatory properties could also be related to the ability to properly anticipate a stimulus. Together, these results strongly implicate the role of alpha oscillations in hemifield-based spatial attention tasks and for prediction of relatively simple stimuli.

One issue with relating the extant literature on alpha oscillations to prediction is that is unclear whether the experiments described therein engaged actual predictive processing about what would happen or comparatively simple anticipatory attention mechanisms about *where* a stimulus might appear. For example, the alpha rhythm might simply correspond to shifts of spatial attention, which can be oriented approximately 10 times per second (VanRullen & Dubois, 2011), and studies using a Posner spatial cueing task (Posner, 1980) support this idea (Capotosto, Babiloni, Romani, & Corbetta, 2009; Busch & VanRullen, 2010). Other studies, however, suggest that prediction and attention are separable mechanisms with dissociable effects (Kok, Rahnev, Jehee, Lau, & de Lange, 2012; Wyart, Nobre, & Summerfield, 2012; Horschig, Jensen, van Schouwenburg, Cools, & Bonnefond, 2013). Experiments have generally not examined the role of alpha oscillations in complex perceptual tasks that require actual predictive processing, although it was recently demonstrated that an alpha-band presentation rate (12.5 Hz) is optimal for maximal fMRI BOLD response in a dynamic face recognition experiment (Schultz, Brockhaus, Bulthoff, & Pilz, 2013).

The work described in this chapter investigated the role of predictive processing during a novel object recognition task. The experiment made use of novel three-dimensional stimuli that required integration over multiple sequential views to extract their three-dimensional structure. The stimuli were presented at central fixation and thus, manipulating the ordering of the views could be used to test the effect of their predictability without confounding it with spatial attention. This is henceforth referred to as "spatially" predictable, although it is likely a combination of spatially- and featurally- predictive processing (i.e., the prediction of specific features at specific locations of a spatial map). To test the relationship between alpha oscillations in predictive processing, the

experiment took advantage of the entrainability of endogenous alpha oscillations by exogenous rhythmic stimulation (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Calderone, Lakatos, Butler, & Castellanos, in press). Whereas previous experiments related pre-stimulus alpha to anticipation via a post-hoc grouping of detected and missed stimuli (Mathewson et al., 2009; Busch et al., 2009), the current experiment entrained alpha oscillations by presenting subsequent views of the stimulus at a regular 10 Hz rate to determine the causal effect of alpha oscillations on prediction (henceforth referred to as "temporally" predictable).

The results of the experiment indicated that spatial and temporal predictability of an entraining sequence enhanced discriminability of a subsequently presented probe stimulus. Temporal predictability also speeded response times for the probe judgement. EEG amplitude averaging indicated separable time courses for spatial and temporal prediction with temporal predictability always preceding the onset of stimuli and spatial predictability manifesting at the onset of stimuli and persisting for over 100 ms after in the case of the probe. Oscillatory analyses indicated strong bilateral alpha power and phase coherence modulation as a function of stimulus predictability, as well as similar and sometimes more prominent, effects in the lower frequency delta-theta (5 Hz) band.

1.2 Methods

1.2.1 Participants

A total of 58 students from the University of Colorado Boulder participated in the experiment (ages 18-28 years, mean=21; 31 male, 27 female). EEG was recorded from 29 of the participants while they completed the experiment. The remaining 29 participants completed a solely behavioral experiment without EEG recording. 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.

1.2.2 Stimuli

Novel "paper clip" objects similar to those used in previous investigations of three-dimensional object recognition (Bulthoff & Edelman, 1992; Edelman & Bulthoff, 1992; Logothetis, Pauls, Bulthoff, & Poggio, 1994; Logothetis, Pauls, & Poggio, 1995; Sinha & Poggio, 1996) 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 (Balas & Sinha, 2009b). 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 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 1.1.

1.2.3 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 initial view of the sequence followed by seven additional views spaced 24 degrees apart (Figure 1.2A, 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 1.2A, magenta tick marks) and from outside of the span of the entraining sequence in increments of 24 degrees (Figure 1.2A, green tick marks).

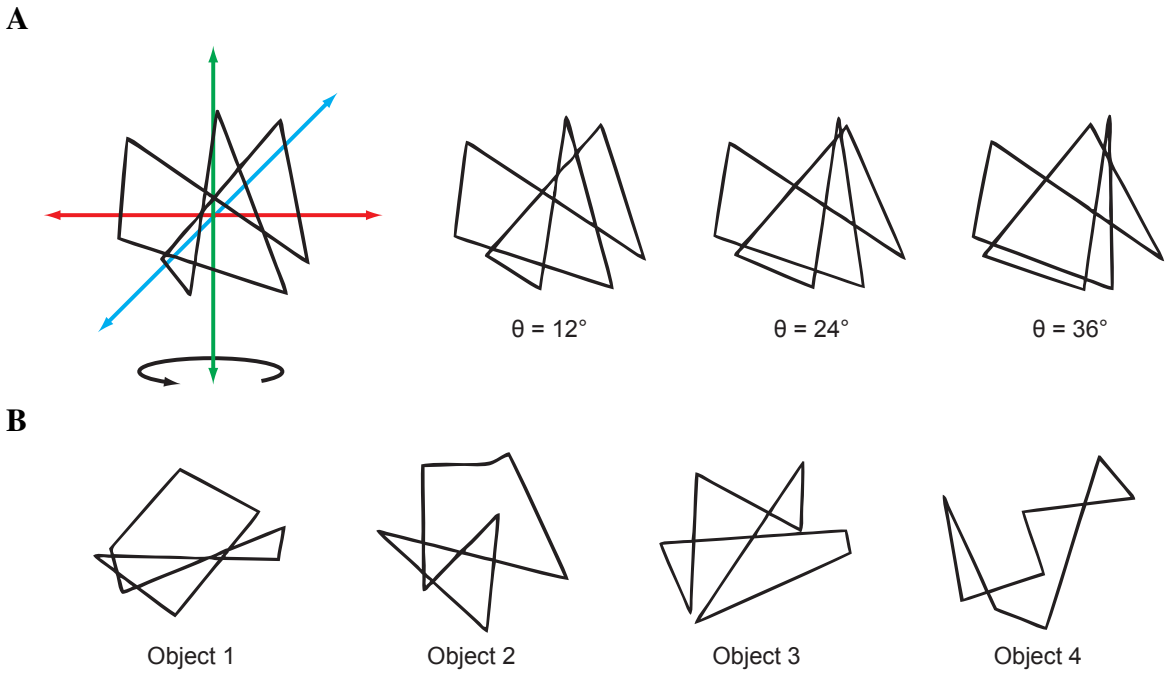

Figure 1.1: Novel "paper clip" objects

A: Objects were composed of eight three-dimensional vertices joined together with line segments. To render the objects to bitmap images, each object was rotated completely about its vertical axis in steps of 12 degrees and reduced to an orthographic projection. **B:** Four of the 16 objects used in the experiment.

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.

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 1.2B). 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.

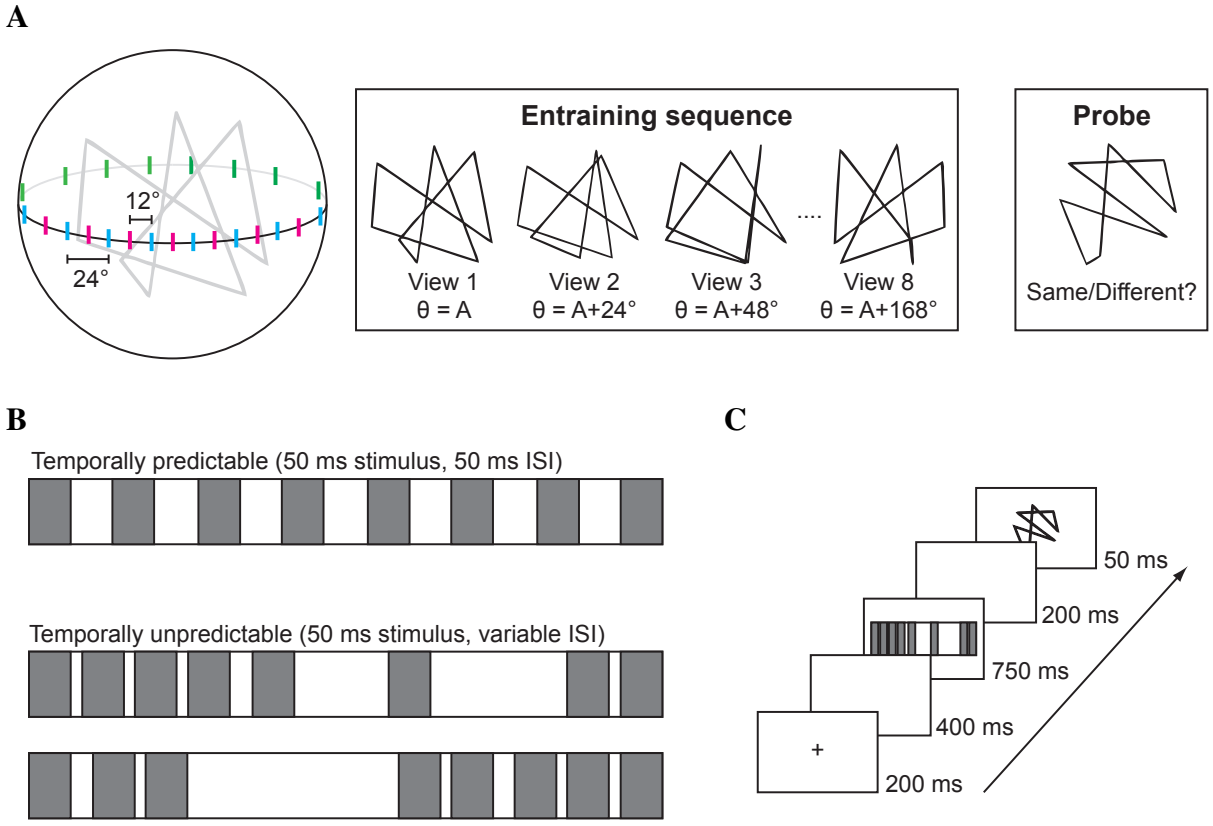

Figure 1.2: Experimental procedure

A: Experimental trials contained an entraining sequence composed of eight views of a single object, followed by a probe stimulus. Entraining views were spaced 24 degrees apart (blue tick marks). The probe depicted an unseen view from the 12 degree interpolations between views of the entraining sequence (magenta tick marks) or from outside the span of the entraining sequence in increments of 24 degrees (green tick marks). **B:** Entraining views were either presented at 10 Hz with a 50 ms on time and 50 ms off time or in a temporally unpredictable manner with a 50 ms on time (gray segments) and variable off time (white segments). In both cases, the duration of the total entraining sequence was held constant at 750 ms. **C:** Order and timing of events within a single trial.

The experiment was displayed on an LCD monitor at native resolution operating at 60 Hz using the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). Stimuli were presented at central fixation on an isoluminant 50% 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 1.2C). 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 participants 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.

1.2.4 EEG recording and preprocessing

The EEG was recorded using an Electrical Geodesics, Inc. (EGI) system composed of a 128 channel net (HCGSN 130) amplified through 200 MΩ amplifiers (Net Amps 200). The signal was sampled at 250 Hz with impedances for each electrode were adjusted to less than 40 kΩ before and during the recording. Stimulus and response trigger onsets were measured via the Psychophysics Toolbox using a high precision realtime clock that was synchronized within 2.5 ms of the EEG system's clock before every trial during the experiment.

EEG data were preprocessed using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). 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 (Delorme & Makeig, 2004) 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 com-

ponents and any bad channels were replaced using spherical spline interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989). The resulting segments were re-referenced to the average reference.

1.2.5 Event-related averaging

Event-related averaging was performed separately for the entraining sequence and the subsequent probe. For the entraining sequence, data were aligned to the onset of entraining views 2 through 8 and averaged from the period beginning 50 ms before each entrainer and ending 50 ms after. Baseline correction was performed using the first 50 ms of this period. For the probe, data were aligned to the probe onset and averaged from the period beginning 200 ms before the probe and ending 400 ms after. This allowed detection of predictability effects during the blank period due to differences in phase elicited by the entraining sequence as well as probe-evoked predictability effects.

All waveforms were averaged over a montage of 23 electrodes that covered the occipital and parietal cortices (Figure 1.3). The montage included locations from the 10-10 system that are commonly associated with perceptual processing (Oz, O1/O2, PO3/PO4, and PO7/PO8) (e.g., Doherty, Rao, Mesulam, & Nobre, 2005; Rohenkohl & Nobre, 2011; Fahrenfort, Scholte, & Lamme, 2007).

1.2.6 Time-frequency analysis

Segmented data were used to compute time-frequency data for each trial. Data were first downsampled to 125 Hz and then used to compute the instantaneous Fourier coefficients at each time bin using a multi-taper approach. Hanning tapers were generated at 5-40 Hz and convolved with the data using a sliding time window (four cycles per frequency per time window). The relatively long time window required for low-frequency bands prevents computation of time-frequency data for short time segments, such as the 200 ms blank before the probe. To address this issue, time-frequency data was computed over the entire 2350 ms trial epoch and then cropped to investigate temporal regions of interest.

Power was computed from the magnitude of the instantaneous Fourier coefficients at each

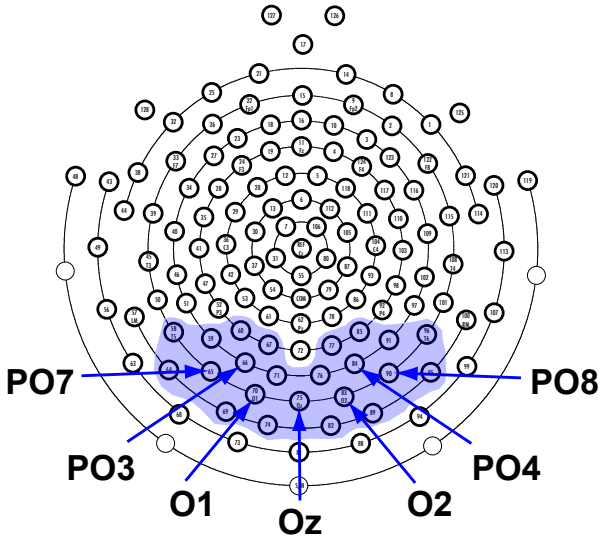

Figure 1.3: Electrode pooling for analyses

Blue shaded region denotes pooled electrodes. Locations from the 10-10 system are indicated.

frequency (f) and time bin (t):w

$$Power(f,t) = \frac{1}{n}\sum_{k=1}^{n}|F_k(f,t)|^2$$

Phase, which is normally defined as the angle of the complex Fourier coefficients, cannot be averaged due to its circularity and thus standard statistical models cannot be applied to assess its significance. One solution to this problem is to compute inter-trial coherence (ITC) instead (Lachaux, Rodriguez, Martinerie, & Varela, 1999). ITC is averaged in the complex domain by first normalizing phase information to unit length by dividing off power and computing the magnitude:

$$ITC(f,t) = \left|\frac{1}{n}\sum_{k=1}^{n}\frac{F_k(f,t)}{|F_k(f,t)|}\right|$$

ITC ranges between 0 and 1 and represents how systematic phase angles are across trials. A value of 0 indicates that phase information is essentially uniformly distributed across trials while a value of 1 indicates a high degree of phase-locking at a particular frequency across trials.

All time-frequency analyses were averaged over the same montage of 23 occipitoparietal electrodes that was used to compute event-related averages (Figure 1.3).

1.3 Results

1.3.1 Behavioral measures of spatial and temporal predictability

Five subjects were excluded from behavioral analysis for accuracy 2.7σ (or further) below mean accuracy across subjects. The remaining 53 subjects were submitted to a 2x2 ANOVA with spatial and temporal predictability as within-subjects factors. Experiment type (EEG or behavioral only) was included as an additional between-subjects factor to ensure that it did not interact with any of the within-subjects factors. Accuracy and reaction times were collected during the experiment and were used to compute *d'*, a measure of sensitivity that takes into account response bias, and inverse efficiency, a measure that combines accuracy and reaction times (Townshend & Ashby, 1978). These behavioral measures are plotted in Figure 1.4.

Subjects that completed the full EEG experiment were on average less accurate ($F(1, 51) = 4.80$, $p = 0.033$) but responded more quickly ($F(1, 51) = 10.05$, $p = 0.003$) than subjects that completed the solely behavioral experiment. These differences reflect a speed-accuracy tradeoff, likely due to differences in instructions given to subjects by experimenters or motivational differences between subject groups. Importantly, experiment type did not interact with any within-subjects factors (all p's > 0.05) indicating that the behavioral measures of interest were not dependent on which type of experiment subjects completed.

Overall, subjects were more accurate when the entraining sequence was temporally predictable ($F(1, 51) = 17.84$, $p < 0.001$). A similar effect for spatial predictability failed to reach significance ($F(1, 51) = 1.85$, $p = 0.18$). The interaction between spatial and temporal predictability, however, was significant ($F(1, 51) = 6.13$, $p = 0.017$). The LeabraTI model (Chapter **??**) as well as investigations predictability on attentional allocation (e.g., Doherty et al., 2005; Rohenkohl, Gould, Pessoa, & Nobre, 2014) suggest that spatial and temporal predictability should have a superadditive effect on behavioral outcomes. However, the combined spatial and temporal predictability condition here (denoted S+T+ in Figure 1.4) was subadditive. Although not significantly different from spatial predictability alone ($t(52) = 1.29$, $p = 0.204$) or from temporal predictability alone

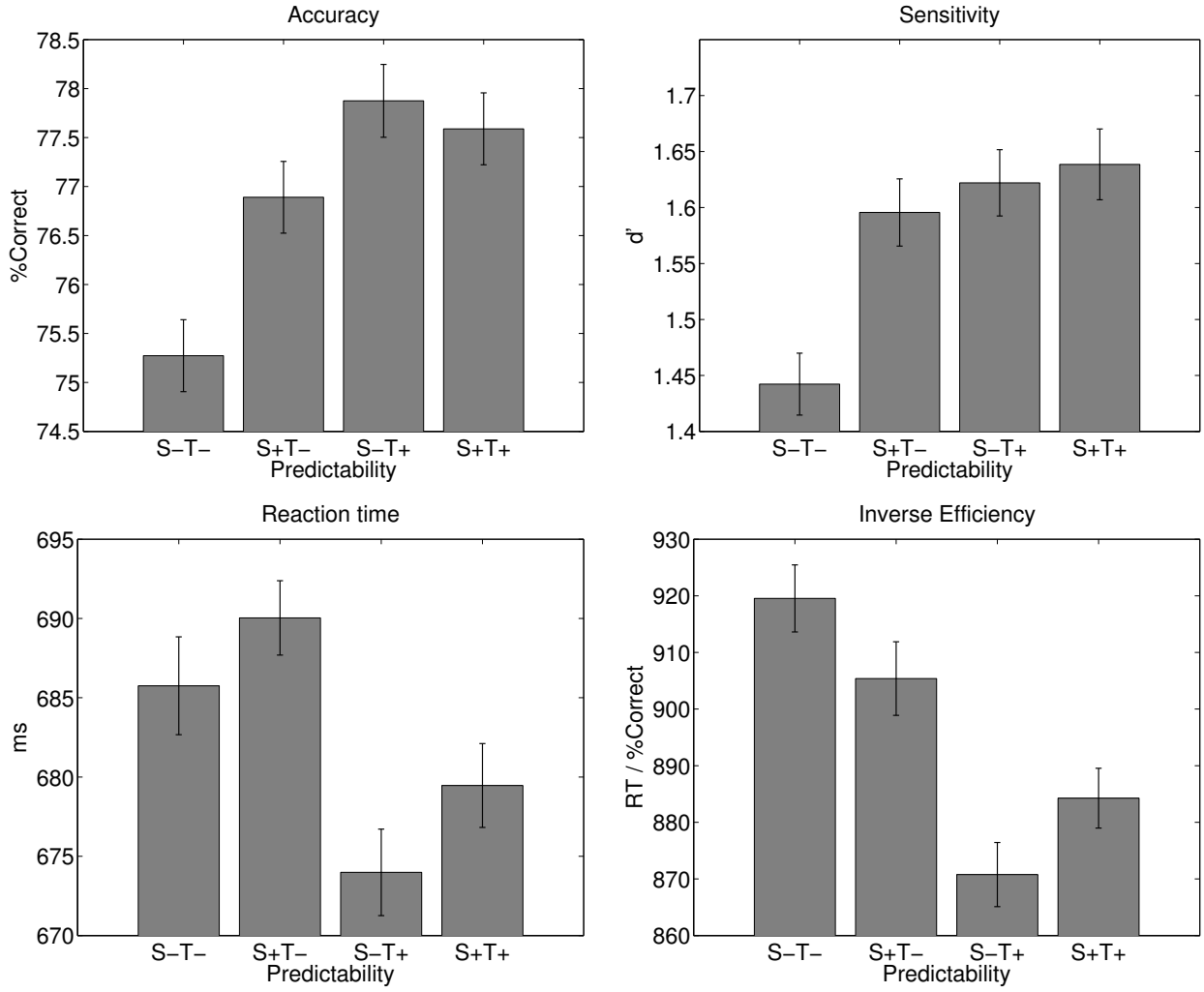

Figure 1.4: Behavioral measures of spatial and temporal predictability

Accuracy, *d'* (sensitivity), reaction time, and inverse efficiency (reaction time divided by percent correct) as a function of entrainment condition. S-/+ refers to spatially unpredictable and predictable, T-/+ to temporally unpredictable and predictable. Error bars depict within-subjects error using the method described in Cousineau (2005) adapted for standard error.

($t(52) = 0.45$, $p = 0.652$), this result merits further investigation.

When responses are transformed into *d'*, there is a significant effect of both spatial ($F(1, 51) = 4.71$, $p = 0.035$) and temporal predictability ($F(1, 51) = 11.99$, $p < 0.001$). This result suggests that response bias can at least partially explain why spatial predictability failed to reach significance for raw accuracy. The interaction between spatial and temporal predictability remained significant

for *d'* ($F(1, 51) = 4.49$, $p = 0.039$). The interaction is additive, but is driven primarily by the strong effect of introducing spatial or temporal predictability over complete unpredictability (S-T- versus S+T-, $t(52) = 3.19$, $p = 0.002$; S-T- versus S+T+, $t(52) = 4.26$, $p < 0.001$) opposed to any synergistic effect of combined spatial and temporal predictability (S+T+ versus S+T-, $t(52) = 0.90$; S+T+ versus S-T+, $t(52) = 0.31$; both p's > 0.05).

Reaction times were significantly faster when the entraining sequence was temporally predictable ($F(1, 51) = 12.38$, $p < 0.001$). A similar effect for spatial predictability failed to reach significance ($F(1, 51) = 1.96$, $p = 0.168$) nor did the interaction term ($F(1, 51) = 0.05$, $p = 0.83$).

Inverse efficiency, which considers reaction time as a function of accuracy (defined as reaction time divided by percent correct) can be thought of as the amount of energy consumed by the system to produce a behavioral outcome (Townshend & Ashby, 1983). It is often used to remove non-monotonicities present in accuracy or reaction times alone, although that effect is not observed here. Nevertheless, it provides another lens under which to interpret the results, and thus it is considered here. Inverse efficiency was significantly lower when the entraining sequence was temporally predictable ($F(1, 51) = 23.31$, $p < 0.001$), but not when it was spatially predictability ($F(1, 51) = 0.002$, $p = 0.963$). Inverse efficiency is characterized by a significant cross-over interaction ($F(1, 51) = 5.85$, $p = 0.019$). Spatial predictability of the entraining sequence produces lowers inverse efficiency over complete unpredictability. Inverse efficiency is lowest on average when stimuli are temporally predictable, but the addition of spatial predictability causes an increase in inverse efficiency.

1.3.2 Time course of spatial and temporal predictability

A total of five subjects were excluded from EEG analysis – three for an overabundance of artifacts in the EEG recording resulting in low trial counts after rejection and two for accuracy 2.7σ (or further) below mean accuracy across subjects (these two subjects were also excluded from behavioral analyses, see preceding section). The remaining 24 subjects were included in all EEG analyses.

A 2x2 ANOVA with spatial and temporal predictability as within-subjects factors was used to assess statistical significance at each time bin of event-related averages. *p*-values were corrected for a maximum false discovery rate (FDR) of 5% using the method described in Benjamini and Yekutieli (2001). Additionally, effects were only considered significant if they persisted for at least 16 ms.

To investigate the build-up of spatial and temporal predictability over the entraining sequence, activity from the second through final entraining views was averaged for each condition (the first entraining view is unpredictable, so it is omitted from the average). The results of this analysis are plotted in Figure 1.5. The first thing worth noting is that a large 10 Hz periodicity is present for the temporally predictable conditions (S-T+ and S+T+), phase-aligned approximately to the onset of each entrainer. Temporally unpredictable entrainers (S-T- and S+T-) are also approximately periodic. The reason for the 10 Hz periodicity in these conditions despite being temporally unpredictable is likely due to the 750 ms constant duration of the entraining sequence regardless of condition (Figure 1.2B). Presenting eight stimuli in 750 ms with a variable ISI is a 10 Hz presentation rate on average. Still, the temporally unpredictable entrainers exhibit markedly weaker amplitude and are approximately 180 degrees out of phase with the the temporally predictable entrainers.

The effect of spatial predictability manifested 26 ms after the onset of the entrainer and persisted for at least another 24 ms (one quarter of the 10 Hz period). Temporal predictability, in contrast, manifested prior to (-38 through -22 ms pre-stimulus) and at the onset of the entrainer (-6 ms pre-stimulus through 18 ms post-stimulus). The effect of temporal predictability appears to be driven by the antiphase relationship between T- and T+ conditions at these time points. Together, these effects demonstrate differential time courses for spatial and temporal predictability.

Spatial predictability was enhanced when stimuli were temporally predictable. This effect is characterized by the significant interaction between spatial and temporal predictability starting 14 ms after the onset of the entrainer and persisting for at least another 36 ms. This result indicates that the brain is more capable of differentiating between spatially coherent and random sequences

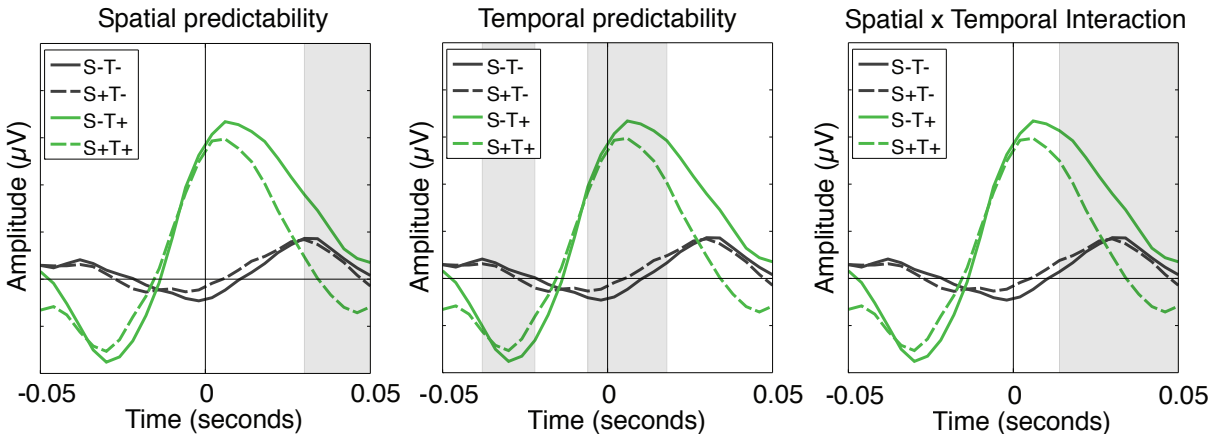

Figure 1.5: Entrainer-evoked activity

Grand averages for entrainers 2 through 8 as a function of entrainment condition. S-/+ refers to spatially unpredictable and predictable, T-/+ to temporally unpredictable and predictable. All plots depict the grand average with gray shaded regions denoting significant effects of spatial predictability (left), temporal predictability (center), and the interaction between these terms controlling for a maximum false discovery rate (FDR) of 5%.

of stimuli when it can properly anticipate the presentation of each stimulus (S-T+ versus S+T+) compared to when the onset is unpredictable.

To investigate the effect of spatial and temporal predictability on perception of the probe, waveforms were aligned to the probe onset onset averaged from 200 ms before through 400 ms after (Figure 1.6). The results of this analysis essentially mirror the entrainer-evoked effects albeit with a few key differences. Temporal predictability was again a purely anticipatory process, manifesting within the pre-stimulus period (-184 ms through -160 ms and -132 ms through -108 ms pre-stimulus). Spatial predictability also manifested briefly within the pre-stimulus period (-128 ms through -80 ms pre-stimulus), which was not seen for spatially predictable entrainers. The post-stimulus effects of spatial predictability occurred much earlier than for entrainers and persisted much longer, beginning approximately at the onset of the probe (-12 ms pre-stimulus) and lasting 136 ms through the P1 response with several transient effects after.

Unlike the significant interaction between spatial and temporal predictability for entrainers, the interaction for the probe failed to reach significance given the constraints of the statistical test

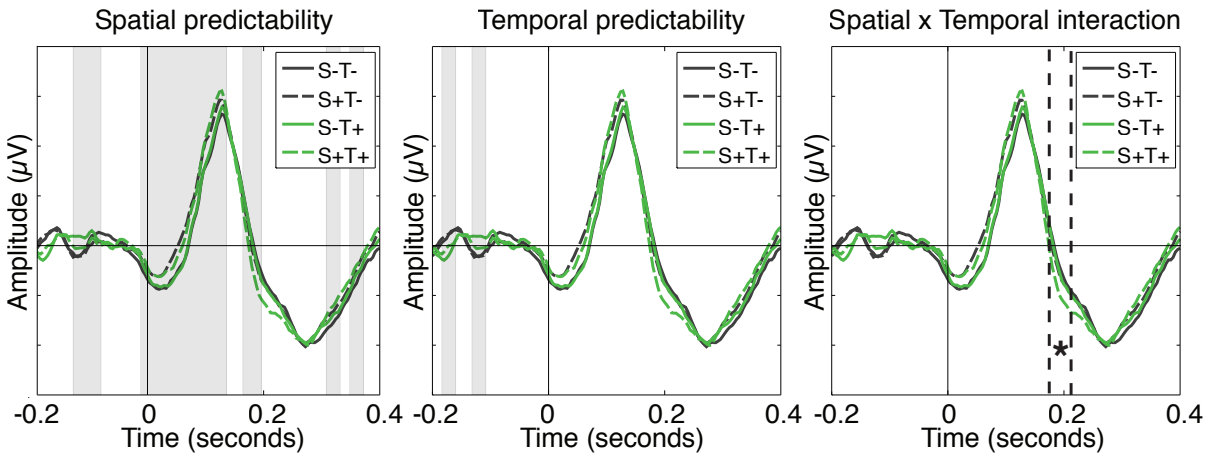

Figure 1.6: Probe-evoked activity

Grand averages for the probe stimulus, including the preceding 200 ms blank. Asterisk in the interaction plot indicates trending significance at the 5% level when averaging amplitude within the window defined by the dotted lines.

(maximum FDR of 5%, 16 ms consecutive significance). However, averaging the amplitude within a window defined by exploratory analysis indicated a trending interaction from 170 ms to 210 ms ($F(1, 22) = 3.97$, $p = 0.059$). This effect was more pronounced and reached significance if only the right hemisphere channels were considered ($F(1, 22) = 5.61$, $p = 0.027$), but failed to reach significance in the left hemisphere ($F(1, 22) = 2.24$, $p = 0.148$), explaining the trending effect when both hemispheres are considered jointly.

1.3.3 Predictability entrains alpha oscillations

The same 24 subjects used in event-related analyses (see preceding section) were used in time-frequency analyses described here. Statistical methods were also identical, consisting of a 2x2 ANOVA with spatial and temporal predictability as within-subjects factors and p-values were corrected for a 5% maximum FDR (Benjamini & Yekutieli, 2001). Effects were only considered significant if they persisted for at least 16 ms.

Power and inter-trial coherence (ITC), a measure of phase angle consistency across trials (Lachaux et al., 1999) were computed over a 5-20 Hz frequency range to investigate the rela-

tionship between entrainer predictability and alpha oscillatory properties (Figures 1.7-1.8). Both spatial and temporal predictability had a significant effect on 10 Hz power and ITC beginning around 200 ms after the onset of the first entrainer (power: 160-168 ms after entrainer 1; ITC: 136-208 ms after entrainer 1). Together, these results indicate that 10 Hz entrainment affects both power and phase alignment and takes around 2-3 events to establish. Spatial predictability had a suppressive effect on 10 Hz power and phase alignment whereas temporal predictability had a positive effect. Any interactions between spatial and temporal predictability failed to reach significant levels altogether.

The effect of spatial predictability only persisted for three entrainers on average (duration ranges for 10 Hz power and ITC: 272-280 ms). Temporal predictability, in contrast, exhibited a sustained effect on 10 Hz power lasting nearly the entire entraining sequence. In the case of ITC, the effect of temporal predictability persisted throughout the blank period that separated the entraining sequence and the probe (Figure 1.9). This result indicates that alpha phase remained more aligned for temporally predictable stimuli even without exogenous entrainment.

Predictability effects on 10 Hz ITC failed to reach significance during the presentation of the probe or the following 400 ms period given the constraints of the statistical test (maximum FDR of 5%, 16 ms consecutive significance). An enhancement in ITC could be observed for the combined predictability (S+T+) condition and trended toward significance when averaged over a 100 ms period beginning 155 ms after the onset of the probe ($F(1, 22) = 4.11$, $p = 0.055$). This effect was more pronounced and reached significance if only the right hemisphere channels were considered ($F(1, 22) = 7.45$, $p = 0.012$), but failed to reach significance in the left hemisphere ($F(1, 22) = 1.47$, $p = 0.237$), explaining the trending effect when both hemispheres are considered jointly.

1.3.4 Predictability effects in delta-theta bands

The effects of spatial and temporal predictability on oscillatory properties during the probe period (-200 ms pre-stimulus through 400 ms after) were investigated in the delta-theta bands,

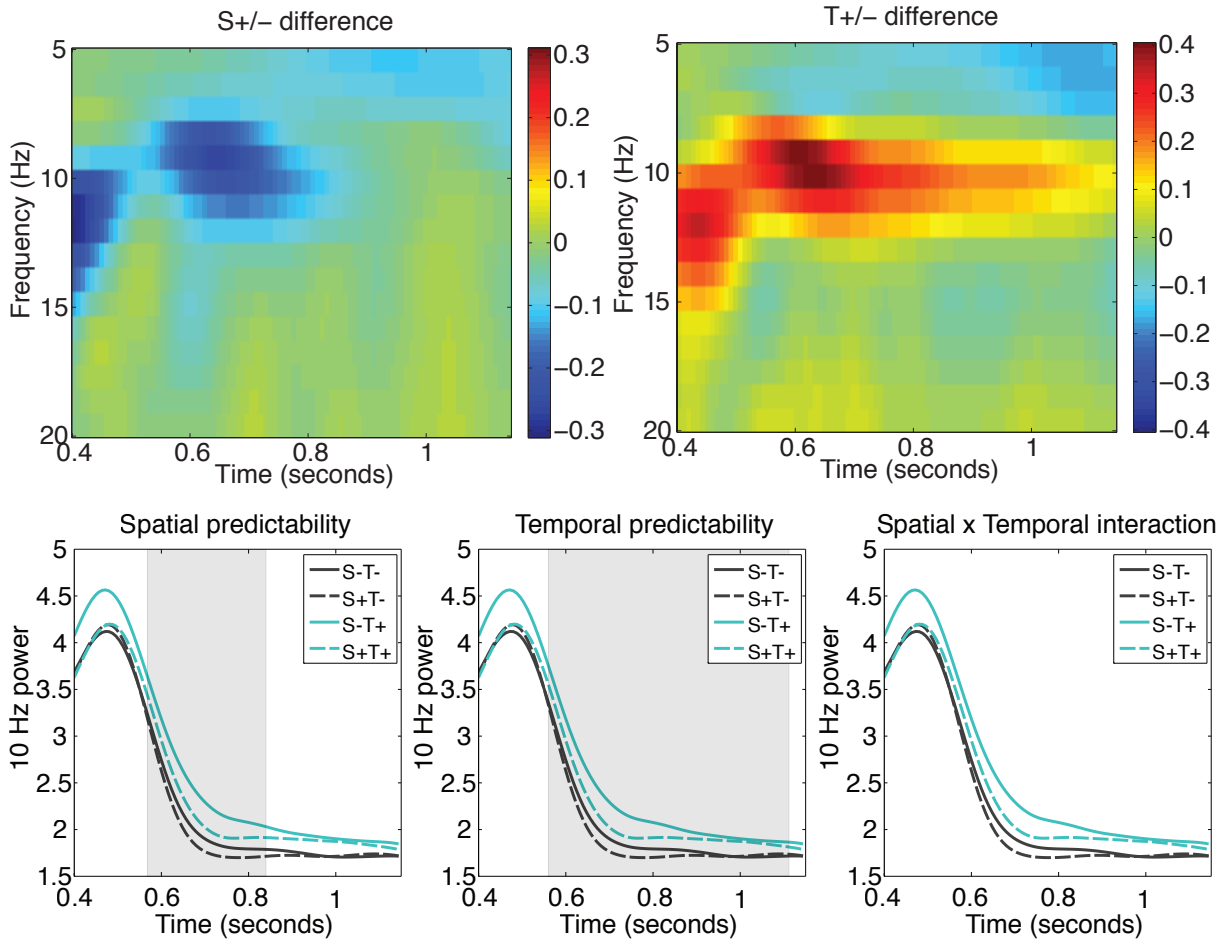

Figure 1.7: Effect of entrainer predictability on alpha power

Alpha-band power over the entraining sequence. **Top:** Main effects of spatial and temporal predictability on oscillatory power in the 5-20 Hz frequency range. **Bottom:** 10 Hz only effects of spatial predictability (left), temporal predictability (center), and the interaction between these terms with gray shaded ranges indicating significance while controlling for a maximum false discovery rate (FDR) of 5%. S-/+ refers to spatially unpredictable and predictable, T-/+ to temporally unpredictable and predictable. Time axes indicate total trial time after the initial fixation cross with 0.4 seconds corresponding to the first entrainer.

centered around 5 Hz. This frequency was identified based on exploratory analysis and was also motivated by alternative models of sensory prediction (e.g., Arnal & Giraud, 2012; Giraud & Poeppel, 2012). Power and ITC at 5 Hz are plotted in Figures 1.10-1.11.

Both spatial and temporal predictability had a significant effect on 5 Hz power during the 200 ms blank period preceding the probe. Temporal predictability had a suppressive effect on 5 Hz

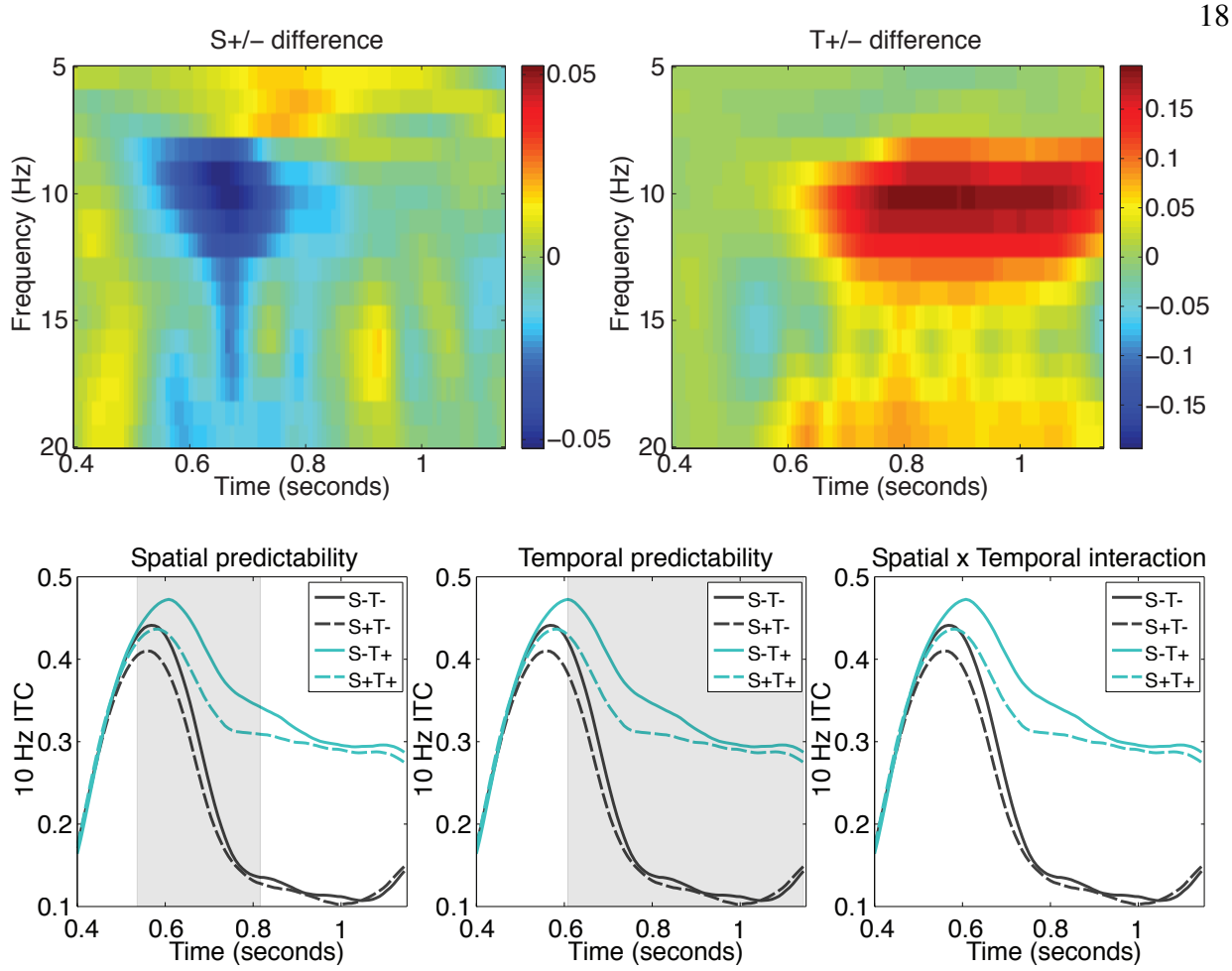

Figure 1.8: Effect of entrainer predictability on alpha phase coherence

Alpha-band inter-trial coherence (ITC) over the entraining sequence. Axes, legends, and shading for significant regions are the same as those described in Figure 1.7.

power, in contrast to the positive modulation found for 10 Hz power. This effect reversed following the presentation of the probe and persisted for over 300 ms. The interaction between spatial and temporal predictability failed to reach significance for 5 Hz power at any time points.

Temporal predictability also had a significant effect on 5 Hz ITC beginning during the 200 ms blank period preceding the probe and lasting nearly 300 ms after the presentation of the probe. Whereas 10 Hz ITC decreased during the blank period (yet remained significantly higher for temporally predictable stimuli), 5 Hz ITC increased, and continued to increase until approximately 100 ms after the onset of the probe. 5 Hz ITC was highest for the combined spatial and temporal

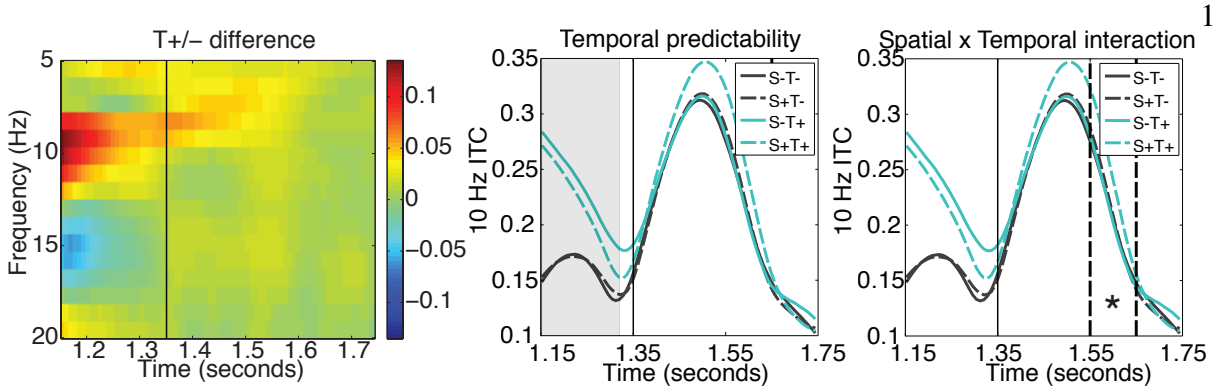

Figure 1.9: Alpha phase coherence before and after probe

Alpha-band inter-trial coherence (ITC) 200 ms preceding the probe and 400 ms following. Solid vertical line indicates probe onset. Asterisk indicates trending significance at the 5% level when averaging 10 Hz ITC within the window defined by the dotted lines. Time axes indicate total trial time after the initial fixation cross.

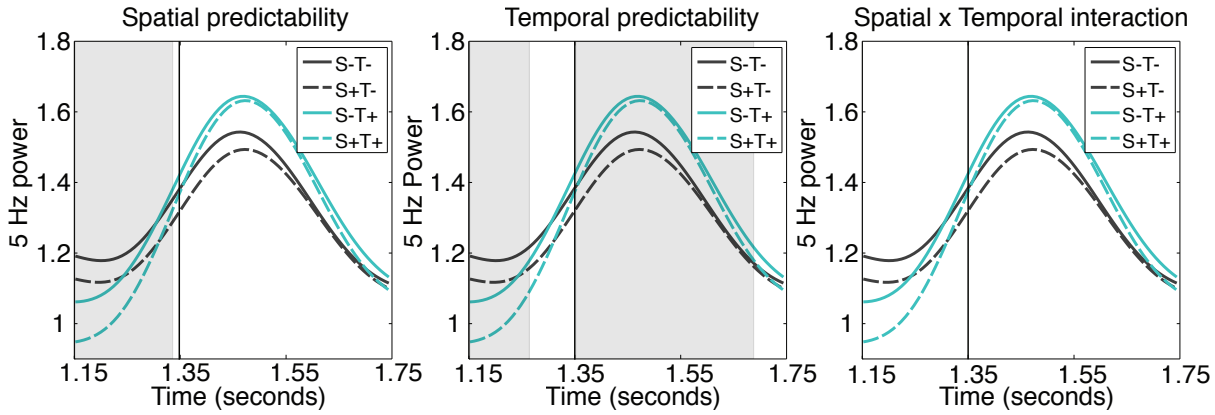

Figure 1.10: Delta-theta power before and after probe

Delta-theta power 200 ms preceding the probe and 400 ms following. Axes, legends, and shading for significant regions are the same as those described in Figure 1.9.

predictability condition (S+T+), indexed by a significant interaction beginning before the probe onset (-78 ms pre-stimulus) and persisting 154 ms after.

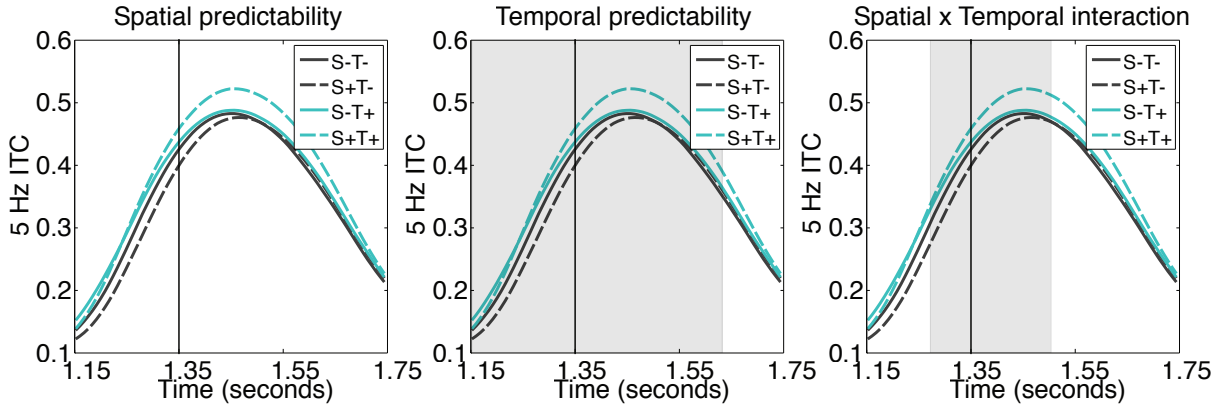

Figure 1.11: Delta-theta phase coherence before and after probe

Delta-theta inter-trial coherence (ITC) 200 ms preceding the probe and 400 ms following. Axes, legends, and shading for significant regions are the same as those described in Figure 1.9.

1.4 Discussion

1.4.1 Summary of results

The work described in this chapter investigated how the brain integrates information from 100 ms samples and uses it to drive predictions about what will happen. The experimental paradigm used to address this question involved entraining alpha oscillatory activity to determine the effects of spatial and temporal predictability on a novel object recognition task. Behaviorally, spatial and temporal predictability increased probe discriminability on a same-different judgement as well as speeded response times. Inverse efficiency, which combines accuracy and reaction times and can be thought of as the amount of energy consumed by the system to produce a behavioral outcome (Townshend & Ashby, 1978, 1983), was lower for temporally predictable stimuli on average, but exhibited an increase for combined spatial and temporal predictability.

Event-related analysis of EEG data indicated that temporal predictability causes a strong periodicity (in this case, 10 Hz) phase aligned approximately to the onset of each temporally predictable entraining stimulus. This alignment approximately 180 degrees out of phase for temporally unpredictable stimuli and these differences in waveform alignment caused amplitude differ-

ences that preceded both entraining stimuli and the probe. The effects of spatial predictability generally manifested during the presentation of stimuli. There was an early divergence of probe-evoked activity caused by spatially predictable ordering of entrainers that over 100 ms through the P1 response with several transient effects after. There was evidence that spatial predictability was only effective when stimuli were also temporally predictable. This was evident for entrainers, and over right hemisphere channels during the probe. The effect failed to reach significance for left hemisphere channels and only trended toward significant levels when both hemispheres were considered jointly. It is likely that this is simply an issue of insufficient power, but other explanations such as hemispheric specialization (e.g., Dien, 2009) cannot be completely ruled out.

Both spatial and temporal predictability had effects on the power and phase coherence (indexed by inter-trial coherence, ITC) of neural oscillations. Spatially predictable entrainment caused a suppression of 10 Hz power with a lower degree of phase alignment than spatially random stimuli. Temporally predictable entrainment had the opposite effect, with increased 10 Hz power and phase alignment. Phase alignment due to temporal predictability remained elevated compared to temporally unpredictable stimuli during a 200 ms blank period between the entraining sequence and probe, indicating phase alignment could persist without exogenous entrainment. There was evidence of selective an enhancement in phase alignment for the combined spatial and temporal predictability case. As was the case in event-related analyses, this effect was significant in the right hemisphere, but not the left, and trended toward significance when both hemispheres were considered jointly. Power and phase coherence effects were also examined in the delta-theta bands (5 Hz) during the probe judgement. Results were similar to those found for 10 Hz, but more robust with effects for both power and phase alignment.

1.4.2 Separate time courses for spatial and temporal prediction

Spatial and temporal predictability were characterized by distinct and generally non-overlapping time courses. Temporal predictability manifested solely before (or at) the onset of each stimulus and appeared to be driven by an approximate antiphase relationship between tem-

porally predictable and unpredictable stimuli. Spatial predictability generally manifested during the presentation of stimuli, although in the case of the probe, showed transient difference without stimulation. Predictability effects were similar for entraining stimuli and the probe except that the effect of spatial predictability persisted for over 100 ms through the P1 response with several more transient effects after. The similarity between entrainers and the probe suggests that the brain might treat the probe as a continuation of the entraining sequence and process it in the same manner.

From these results, it can be concluded that temporal prediction is an anticipatory process, occurring during the absence of exogenous stimulation (in between entrainers or before the probe). The effect of temporal predictability extends 16 ms after the onset of each entrainer, but latencies for the first wave of responses in primary visual cortex (V1) are approximately 40-60 ms (Nowak & Bullier, 1997; Foxe & Simpson, 2002) so there is no exogenous stimulation *per se* during the duration of the effect despite the stimulus being onscreen. Spatial prediction begins shortly before exogenous stimulation, but in the case of the probe, persists through the initial V1 responses. Spatial prediction, thus, might better be characterized as a post-stimulus process opposed to truly anticipatory process. The computation might involve a comparison between what is expected and what is actually coded by incoming spikes, consistent with the LeabraTI model (Chapter **??**) as well as predictive coding models (e.g., Rao & Ballard, 1999).

1.4.3 Oscillatory mechanisms of spatial and temporal prediction

Spatial and temporal predictability had effects on both power and phase coherence of neural oscillations. In both cases, predictability took 2-3 entraining stimuli to establish, consistent with previous investigations (Mathewson, Fabiani, Gratton, Beck, & Lleras, 2010; Mathewson et al., 2012). Spatial predictability was then characterized by a suppression of 10 Hz power and a phase angle variability causing a lower degree of alignment than spatially random stimuli. This effect is opposite than that of temporal predictability, which was characterized by increased 10 Hz power and phase angle alignment. Previous investigations have generally not simultaneously manipulated temporal rhythmicity and spatial coherence (although see Doherty et al., 2005; Rohenkohl et al.,

2014) and thus, the relative suppression and decreased phase coherence during spatially predictable entrainment were unexpected results.

Successful oscillatory entrainment is thought to be a result of repeated phase resetting of endogenous oscillations causing phase to move into alignment with the frequency of exogenous stimulation (Schroeder et al., 2008; Calderone et al., in press). Oscillatory phase resetting has been shown to be caused by salient, unexpected events (Fiebelkorn et al., 2011; Landau & Fries, 2012; Romei, Gross, & Thut, 2012) and thus these unexpected results might be accounted for by considering successive views in terms of the amount of "surprise" (Itti & Baldi, 2009; Meyer & Olson, 2011) they evoke. Subsequent views of the entraining sequence have significant feature overlap, characterized by the same populations of neurons spiking from one view to the next. Repeated spiking, especially as a function of expectation, has been shown to evoke rate suppression mechanisms (Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008). When entraining views are presented out of order, feature overlap is minimized and each view is "surprising", causing an initial fast spiking burst response, which could lead to a higher degree of phase resetting.

The entrainment effects of spatial predictability were transient and dissipated before the end of the entraining sequence whereas the effects of temporal predictability persisted much longer. This might account for the null effects of spatial predictability on most behavioral measures. Temporal predictability effects persisted through the 200 ms blank period between the entraining sequence and probe and had robust effects on all behavioral measures. Thus, the present experiment could be modified to use a shorter entrainment sequence which would likely elicit successful spatial predictability for probe judgements.

The enhancement of 10 Hz phase angle alignment after the probe was presented did not reach significant levels assuming the FDR-corrected significance test at each time bin but did for 5 Hz phase alignment. Furthermore, temporal and spatial predictability main effects around the probe onset were more pronounced for 5 Hz oscillatory properties. One potential explanation for these effects is that reason for this is that the 200 ms blank period between the entraining sequence and the probe corresponded to two cycles at 10 Hz, but only one cycle at 5 Hz. Phase angles

at 10 Hz exhibited significant dealignment over this period, and actually increased in alignment at 5 Hz. Thus, it is possible that this increase in phase alignment lead to a more pronounced selective enhancement for combined spatial predictability at 5 Hz, consistent with recent data (Cravo, Rohenkohl, Wyart, & Nobre, 2013). A more robust effect might be found for 10 Hz if the blank period between the entraining sequence and probe was only 100 ms in duration.

Another limitation of the present experimental paradigm is that it is unclear whether exogenous entrainment simply created new oscillations akin to steady state visually evoked potentials (SSVEP), or actually entrained existing endogenous oscillations. Thus, it might not be particularly surprising that temporally predictable stimuli caused increases in power and phase alignment. However, entrained alpha-band periodicity has been shown to correlate with individual resting alpha oscillation frequency (de Graaf et al., 2013) so it is likely that the paradigm recruited existing oscillations and caused them to align to the exogenous entrainment frequency. The fact that phase alignment continued through the 200 ms blank period without exogenous entrainment also supports this claim.

1.4.4 Alpha oscillations index stimulus-predictive processing

Alpha oscillations have previously been implicated in the allocation of hemifield-based spatial attention and anticipation of the temporal onset of relatively simple stimuli (Gould et al., 2011; Belyusar et al., 2013; Rohenkohl & Nobre, 2011; Mathewson et al., 2009; Busch et al., 2009). It was unclear whether these results engaged actual predictive processing about what would happen or comparatively simple anticipatory attention mechanisms about *where* a stimulus might appear.

The current experiment consisted of a relatively complex perceptual tasks that that required from integration across features extracted over the course of a sequence of stimulus presentations to perform a subsequent probe judgment. Both spatial and temporal predictability benefited probe discriminability and so it can reasonably be concluded that the experiment engaged actual predictive processing. Alpha oscillations indexed both the spatial and temporal predictability of the entraining stimuli as well as tracked the onset of the probe through phase alignment during the 200

ms blank period separating the entraining sequence and probe.

Cumulatively, these results suggest that alpha oscillations serve at least two roles in prediction that can roughly be characterized as "spatial" and "temporal" in nature. The first is a prediction about the spatio-featural content of upcoming sensory events. This prediction could simply be about where a stimulus might show up, irrespective of what features it contains. Importantly however, the prediction can also carry content about the probability of specific features showing up at a specific location in spatial map(Kok et al., 2012; Wyart et al., 2012; Horschig et al., 2013). The second function of alpha oscillations is a pacemaker function that allows predictions to be made at regular intervals. This pacemaker function can phase align to the onset of stimuli (Schroeder et al., 2008; Calderone et al., in press) so that predictions can reliably be made in advance of actual sensory events. Overall, these suggested roles of alpha oscillations are compatible with the LeabraTI model (Chapter **??** and other models of sensory predictions that highlight the role of oscillations in making about making predictions about incoming sensory information (Arnal & Giraud, 2012; Giraud & Poeppel, 2012).

Chapter 2

Effects of spatial and temporal prediction during prolonged learning of novel objects

2.1 Introduction

The core challenge of object recognition is concerned with solving the invariance problem (DiCarlo, Zoccolan, & Rust, 2012). Essentially, object identity must remain be invariant across large changes in a an object's visual position, scale, rotation, and viewpoint to generate successful behavior. Understanding how exactly the brain solves this problem has been a major focus of the object recognition literature with the bulk of data and models suggesting that it is solved gradually by a hierarchy of neural processing mechanisms from V1 through inferior temporal (IT) cortex that extract increasingly complex features at each stage with increasing tolerance to transformations (Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; O'Reilly, Wyatte, Herd, Mingus, & Jilk, 2013).

One question that remains to be fully answered is how invariance is learned in the first place. One intriguing hypothesis is that a temporal association rule might form associations between multiple samples of a single object as it undergoes transformations (Stringer et al., 2006; Wallis & Baddeley, 1997; Isik et al., 2012). It has been demonstrated that some neurons can form temporal associations between arbitrary pairs of stimuli (Sakai & Miyashita, 1991), including a population in monkey IT cortex (Meyer & Olson, 2011). Experiments by DiCarlo and colleagues have indicated that these temporal associations can build new invariance for specific object transformations including changes in position and size (Cox, Meier, Oertelt, & DiCarlo, 2005; Li & DiCarlo, 2008, 2010). This new invariance was learned without supervised reward, suggesting that it could be a

natural consequence of generic neural processing mechanisms given the spatiotemporal statistics of the physical world (Li & Dicarlo, 2012).

Evidence of invariance due to temporal associations has yet to be demonstrated in IT neurons for three-dimensional changes in viewpoint (although see Wallis & Bulthoff, 2001; Wallis, Backus, Langer, Huebner, & Bulthoff, 2009, for relevant human behavioral work). IT neurons typically have a tuning curve of approximately 90 degrees for newly acquired three-dimensional objects (Logothetis et al., 1994; Logothetis et al., 1995), but recognition is possible in a viewpoint invariant manner especially after prolonged learning (Wallis & Bulthoff, 1999). Intuitively, predictable motion from one moment to the next could be considered important for encoding three-dimensional objects (Lawson, Humphreys, & Watson, 1994; Stone, 1998; Vuong & Tarr, 2004; Balas & Sinha, 2009b, 2009a; Chuang, Vuong, & Bulthoff, 2012), and thus a temporal association rule could plausibly be used to learn viewpoint invariance from naturalistic spatial structure of objects.

The work described in this chapter investigated the role of predictable spatiotemporal information during a novel object learning task. In the context of the LeabraTI model (Chapter **??**) as well as several other theories of sensory prediction (Arnal & Giraud, 2012; Giraud & Poeppel, 2012), spatial structure might be learned from predictions about incoming sensory information made at regular temporal intervals. To test this hypothesis, both the spatial and temporal predictability of changes in objects' viewpoint were manipulated during a training period followed by a series of same-different judgements over static test views.

Somewhat surprisingly, the results of the experiment indicated that accuracy was lowest when stimuli were learned in a combined spatially and temporally predictable context and highest when learned in a completely unpredictable context. Reaction times were also slower when objects were learned with spatial predictability.

2.2 Methods

2.2.1 Participants

A total of 62 students from the University of Colorado Boulder participated in the experiment (ages 18-22, mean=19.11 years; 30 male, 32 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.

2.2.2 Stimuli

Novel "paper clip" objects similar to those used in previous investigations of three-dimensional object recognition (Bulthoff & Edelman, 1992; Edelman & Bulthoff, 1992; Logothetis et al., 1994; Logothetis et al., 1995; Sinha & Poggio, 1996) were used as stimuli (see Chapter 1 Methods). A total of eight objects were used – four as targets and four as distractors. The four target objects were also used in the Chapter 1 experiment. Target and distractor objects were paired together for the purposes of the experiment. All objects are shown in Figure 2.1.

2.2.3 Procedure

The experiment was divided into 16 blocks, each containing a training period followed by a series of test trials (Figure 2.2). During the training period of a given block, participants observed one of the target objects rotate about its y-axis. The object either rotated coherently (i.e., spatially predictable, S+ conditions) or in a random manner (S- conditions). Coherent rotation was composed of adjacent views spaced 12 degrees apart. The object made four complete rotations during the study period. All views of the object were still presented four times each in the random case. The presentation rate during the study period was either 10 Hz with a 50 ms on time and 50 ms off time (i.e., temporally predictable, T+ conditions) or variable with a 50 ms on time and off times ranging from 16.67-400 ms (T- conditions).

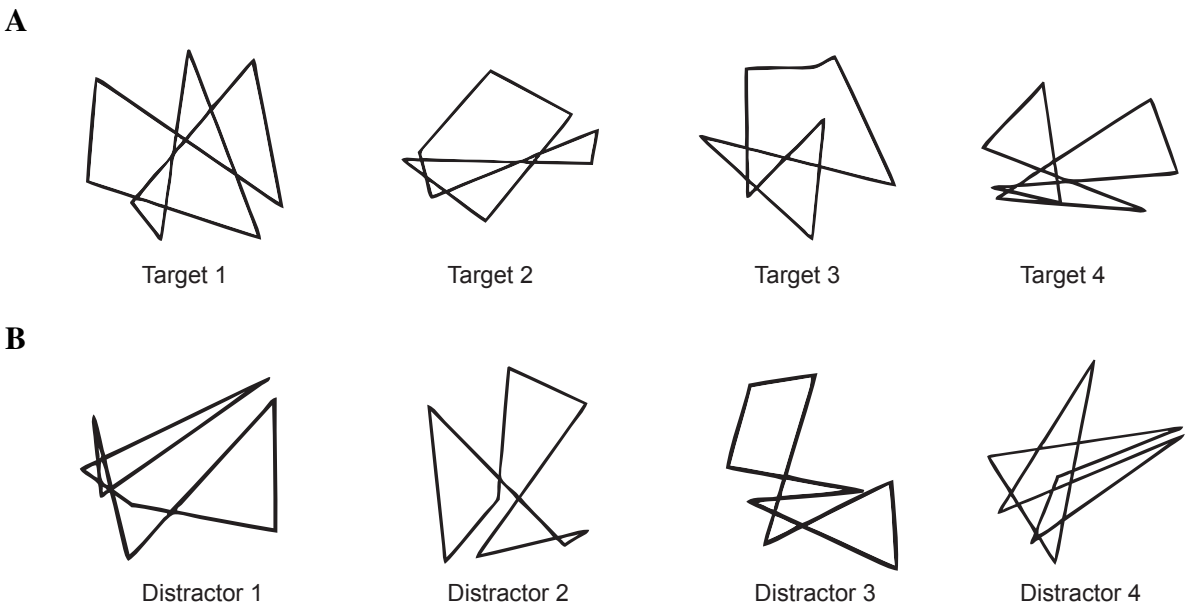

Figure 2.1: Novel "paper clip" objects

Four target (**A**) and four distractor object pairs (**B**) used in the experiment. See Chapter 1 Methods for additional information.

The S+/- and T+/- conditions were crossed and each of the target-distractor object pairs was assigned to one of the four conditions. These assignments were approximately counterbalanced across participants (Assignment 1: N=15; Assignment 2: N=17; Assignment 3: N=15; Assignment 4: N=15). Each block condition with its target-distractor pairing was repeated for four blocks during the experiment. Block order randomized was randomized for each participant.

During each block, participants were instructed to study the target object during the training period and then complete a series of 30 test trials. On each test trial, either the target object or its paired distractor was presented. Participants were instructed to respond "same" if they believed the object depicted the trained target object or "different" if they believed it depicted the distractor object. Half of the test trials contained 15 views of the target object spaced 24 degrees apart, and the other half contained 15 views of the distractor, also spaced 24 degrees apart. Test trials were shown in a random order and feedback was withheld to prevent participants from changing their response criteria over the course of a block.

The experiment was displayed on an LCD monitor at native resolution operating at 60 Hz using the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997). All stimuli were presented at central fixation on an isoluminant 50% gray background and subtended approximately 5 degrees of visual angle. Test trials began with a fixation cross (200 ms) followed by a blank (400 ms) followed by the probe stimulus (100 ms). Participants were required to respond within 2000 ms. Subsquent test trials were separated by a variable intertrial interval of 1000-1400 ms.

The experiment began with a practice block to ensure that participants understood the task. The training period during the practice block was always spatially and temporally predictable and used a reserved target object and distractor that were not further used in any of the experimental blocks. During the practice test trials, participants received feedback after responding according to whether they were correct or incorrect. After completing the practice block, participants were informed that future training periods could be presented in spatially and/or temporally unpredictable manners.

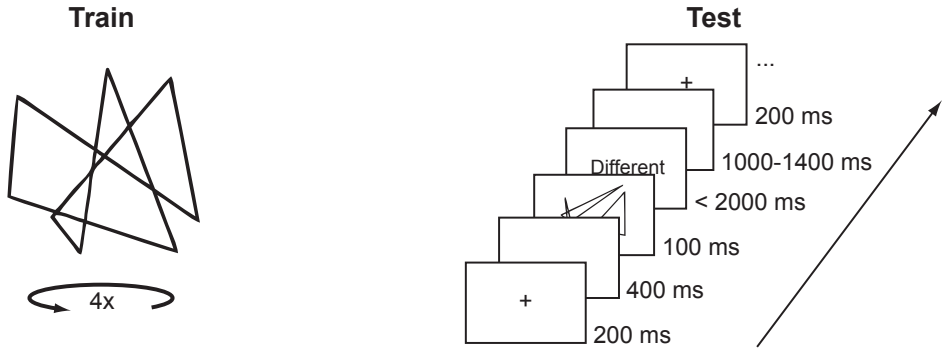

Figure 2.2: Experimental procedure

Experimental trials were composed of a training period followed by a testing period. The training period depicted a target object rotating a total of four times around its vertical axis. Rotation was either spatially and temporally predictable, spatially predictable or temporally predictable only, or completely unpredictable. The test period contained 30 trials that depicted either the training object or its paired distractor at 15 viewing angles each.

2.3 Results

Three subjects were excluded from behavioral analysis for accuracy 2.7σ (or further) below mean accuracy across subjects. All three excluded subjects were assigned condition-object 3 resulting in the final counterbalancing – Assignment 1: N=15; Assignment 2: N=14; Assignment 3: N=15; Assignment 4: N=15. The remaining 59 subjects were submitted to a 2x2 ANOVA with spatial and temporal predictability as within-subjects factors and counterbalancing assignment as a between-subjects factor. Accuracy and reaction times were collected during the experiment and were used to compute *d'*, a measure of sensitivity that takes into account response bias, and inverse efficiency, a measure that combines accuracy and reaction times (Townshend & Ashby, 1978, 1983). These behavioral measures are plotted in Figure 2.3.

Overall, subjects were less accurate when the training period was spatially predictable (F(1, 57) = 4.50, p = 0.038) or temporally predictable (F(1, 57) = 4.20, p = 0.046). The interaction between spatial and temporal predicability failed to reach significance (F(1, 57) = 0.20, p = 0.659). Subjects were least accurate for the combined spatial and temporal predictability condition (denoted S+T+ in Figure 2.3). This condition significantly differed from the completely unpredictable condition (S-T-) (t(58) = -2.8587, p = 0.001), and trended toward significance for conditions with only spatial or only temporal predictability (S+T+ versus S+T-, t(58) = -1.60, p = 0.116; S-T- versus S+T+ versus S-T+, t(58) = -1.77, p = 0.082).

When responses are transformed into *d'*, effects of spatial predictability and temporal predictability during the training period trended toward significance (spatial, F(1, 57) = 3.07, p = 0.085; temporal, F(1, 57) = 3.00, p = 0.089). The interaction between spatial and temporal predicability failed to reach significance (F(1, 57) = 0.00, p = 0.985). The pattern of results as a function of predictability during the training period was the same as for accuracy, and thus this failure to reach critical significance likely reflects the loss of power when transforming responses into *d'* due to discarding misses and correct rejections.

Subjects were slower to respond when the training period was spatially predictable (F(1,

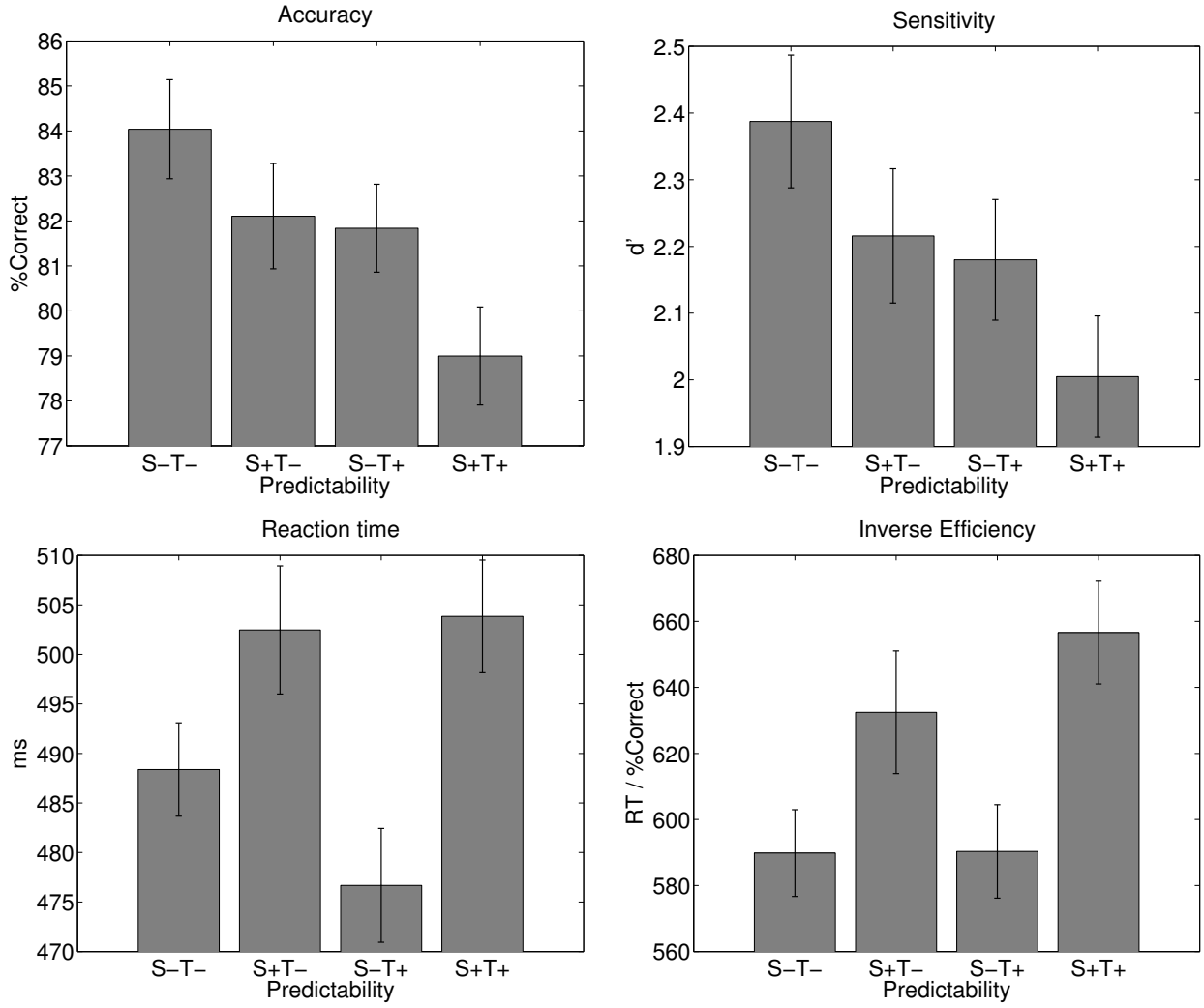

Figure 2.3: Behavioral measures of spatial and temporal predictability

Accuracy, *d'* (sensitivity), reaction time, and inverse efficiency (reaction time divided by percent correct) as a function of predictability during the training period. S-/+ refers to spatially unpredictable and predictable, T-/+ to temporally unpredictable and predictable. Error bars depict within-subjects error using the method described in Cousineau (2005) adapted for standard error.

57) = 10.99, p = 0.002). A similar effect for temporal predictability failed to reach significance (F(1, 57) = 0.53, p = 0.471), nor did the interaction between spatial and temporal predictability (F(1, 57) = 1.21, p = 0.276). Effects on inverse efficiency (defined as reaction time divided by percent correct) were similar. Inverse efficiency was highest when the training period was spatially predictable (F(1, 57) = 9.64, p = 0.003), but did not significantly differ as a function of temporal

predictability ($F(1, 57) = 0.45$, $p = 0.507$), nor when considering the interaction between spatial and temporal predictability ($F(1, 57) = 0.71$, $p = 0.403$).

Effects were highly variable across target objects (Figure 2.4). Target-condition assignment did not significantly affect any of the behavioral measures (all p's > 0.05), but often interacted with predictability effects and their interactions. One reason for this variability regards the orthographic projection used to render the objects. Previous research has indicated that recognition accuracy fluctuates as a function of how well the two-dimensional projection of an object captures its full three-dimensional structure (Balas & Sinha, 2009b). For example, when there is a large amount of foreshortening in the projection, it could be difficult to infer the length of line segments that compose the object, impairing recognition. These degenerate projections are generally diametrically opposed on the object.

To investigate this hypothesis, accuracy was computed as a function of viewing angle for each target object to investigate whether it interacted with predictability during the training period (Figure 2.5). Only accuracy was considered in this analysis as each data point only corresponded to four trials per subject and thus transformation to *d'* was not plausible. Test trials during which distractor objects were presented were also excluded from this analysis since there is no consistent relationship between the targets and distractors across viewing angles and thus they would only contribute noise. With the exception of target object 1, all objects indicated fluctuations in accuracy as a function of viewing angle with two diametrically opposed degenerate views. The most consistent differences in accuracy between training conditions appeared to be localized to the troughs of the accuracy function, corresponding to these degenerate views.

Standard statistical tests did not have enough power to detect differences between conditions for degenerate views due to the low trial counts for each data point. To address this design limitation, a bootstrapping method was used to resample the available data in these cases. The completely unpredictable (S-T-) and combined spatial and temporal predictability (S+T+) were used to assess differences due to training context since these two conditions elicited the greatest difference in average accuracy in the full analysis. The accuracy function over viewing angles was

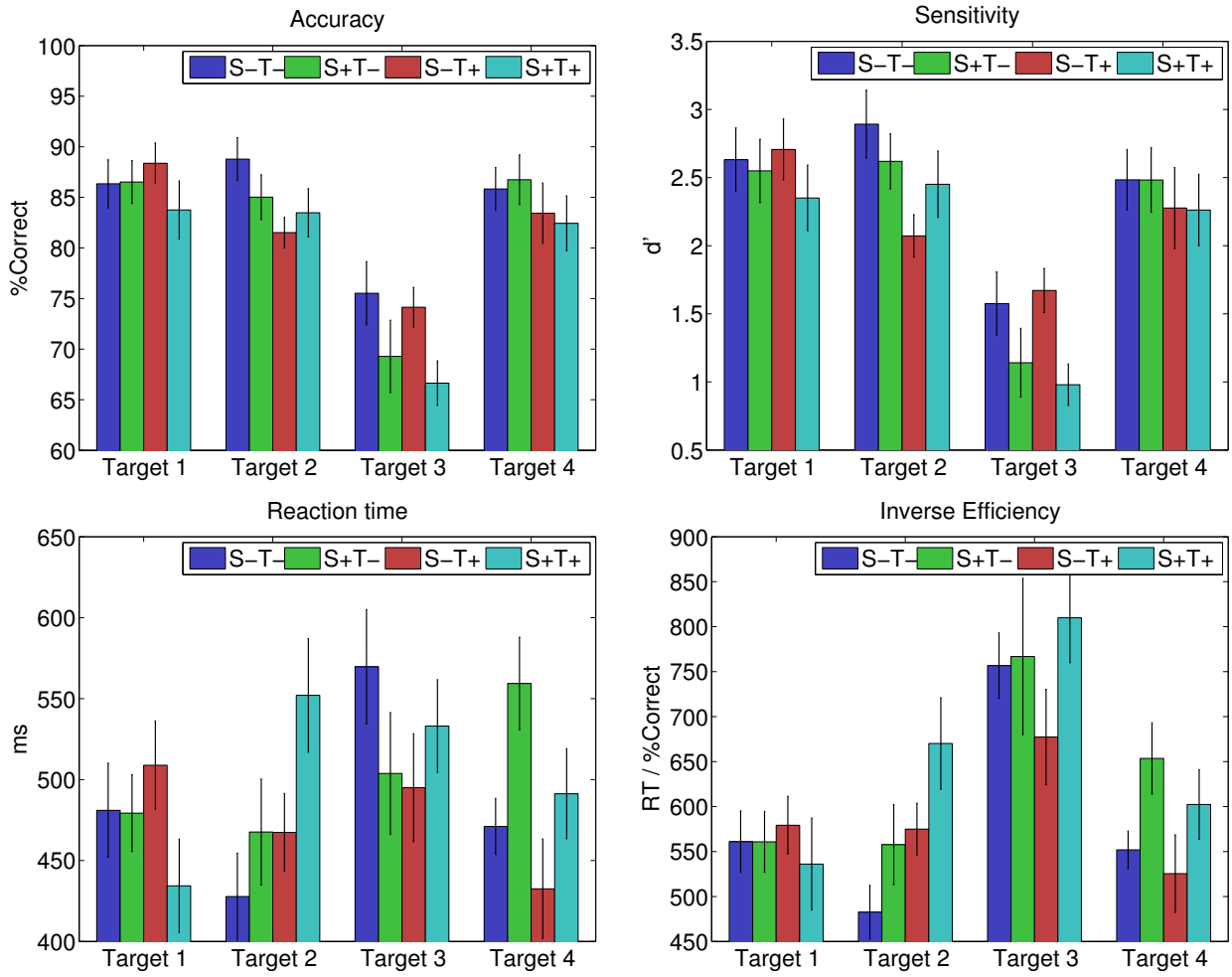

Figure 2.4: Behavioral measures for each target object

Accuracy, *d'*, reaction time, and inverse efficiency for each target object. Horizontal axes denote target and colors predictability during the training period. Error bars depict between-subjects standard error.

collapsed across conditions and the two minima associated with degenerate views were identified for each object. For target object 1, the two views were at $\theta = \{24°, 312°\}$, object 2: $\theta = \{48°, 240°\}$ object 3: $\theta = \{144°, 312°\}$, and object 4 $\theta = \{24°, 192°\}$. S-T- and S+T+ accuracy was for each object's degenerate views and resampled with replacement from the 59 subjects for 10000 iterations. This produced degenerate view accuracy distributions for spatiotemporally unpredictable and predictable training contexts (Figure 2.6). Accuracy for was lower for degenerate views for the spatiotemporally predictable condition for all target objects except target 1, which didn't exhibit

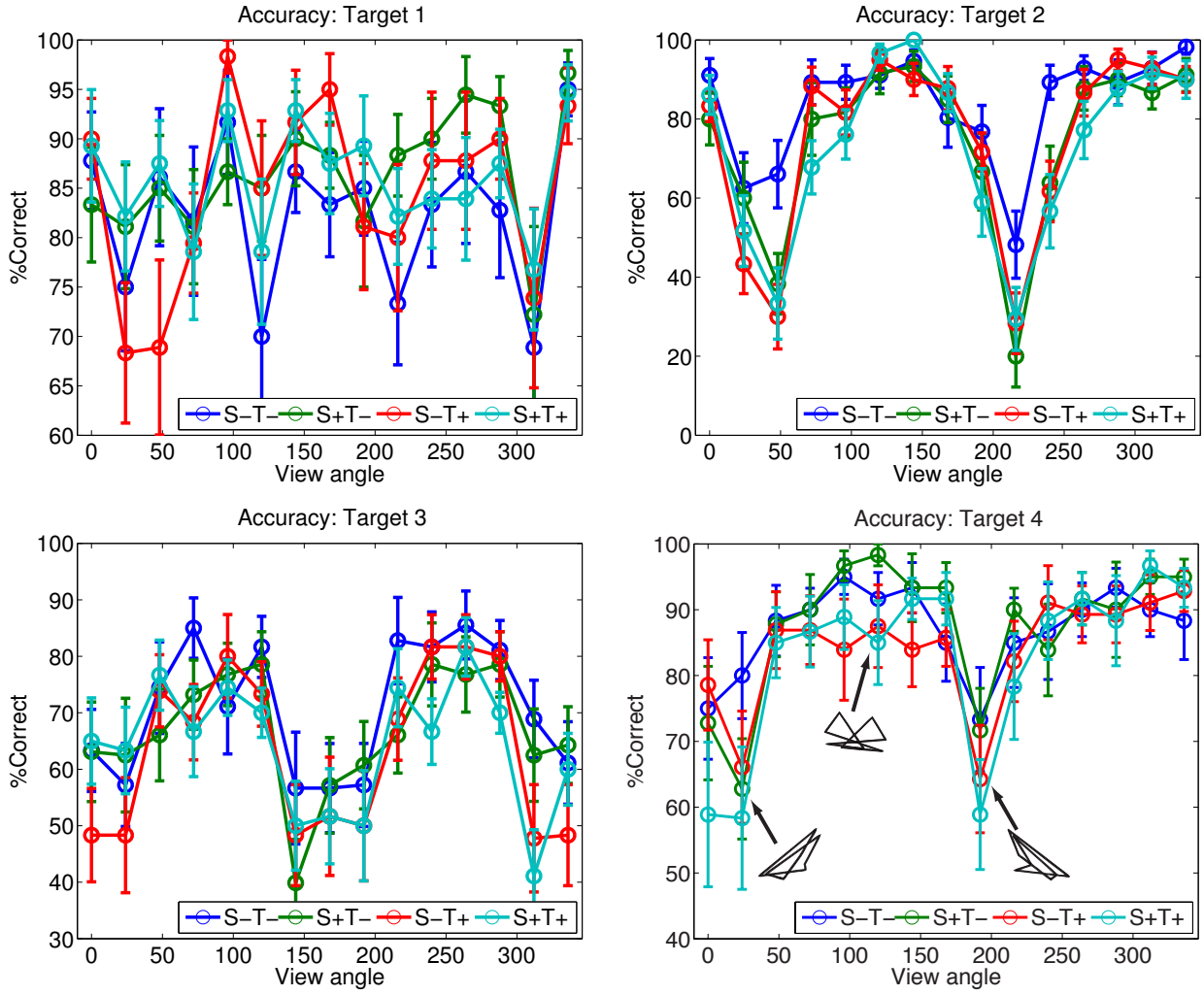

Figure 2.5: Accuracy as a function of viewing angle for each target object

Target accuracy at each viewing angle presented during the test periods. Horizontal axes denote viewing angle and colors predictability during the training period. Error bars depict between-subjects standard error. Diametrically opposed foreshortened views and one canonical view are shown for target object 4.

the patterned accuracy function that other targets did. The predictability difference in accuracy for degenerate views was significant at the 90% alpha level (i.e., the confidence interval of the difference between means did not include zero) for all target objects except target 1.

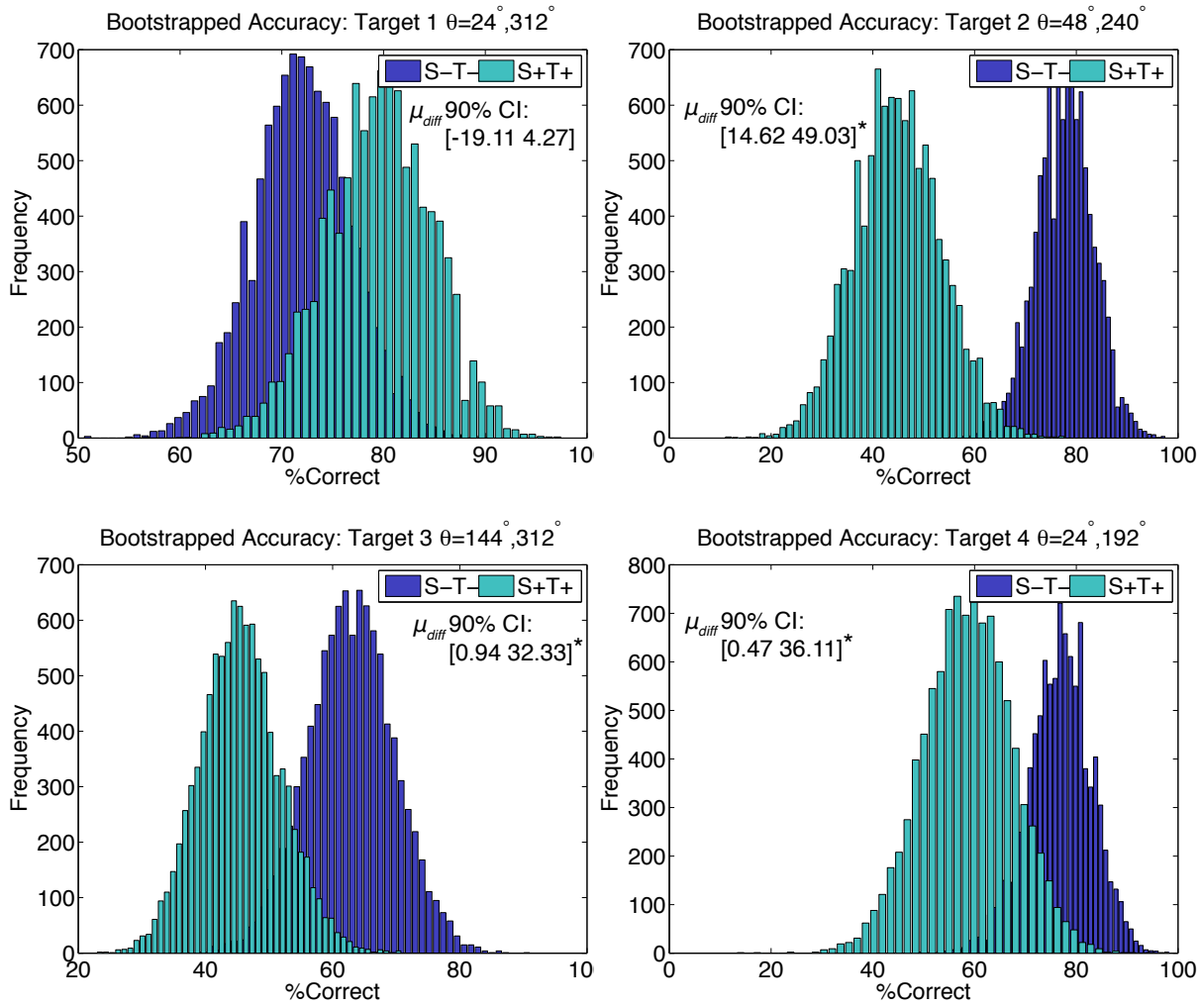

Figure 2.6: Bootstrapped accuracy for degenerate views

Average target accuracy for degenerate views resampled with replacement from the 59 subjects for 10000 iterations. Viewing angle for averaging is noted for each target object. Asterisks denote significant differences based on 90% confidence intervals.

2.4 Discussion

2.4.1 Summary of results

The work described in this chapter investigated how predictability biased learned representations of novel objects. The experimental paradigm used to address this question involved training participants to recognize novel objects while manipulating their spatial and temporal predictabil-

ity. Somewhat surprisingly, accuracy was lowest when stimuli were learned in a combined spatially and temporally predictable context and highest when learned in a completely unpredictable context. Reaction times were also slower when objects were learned with spatial predictability.

Behavioral measures were highly variable across objects. There was some indication that the principal differences between predictability conditions during training were driven primarily by degenerate viewing angles caused by three-dimensional foreshortening in the objects used. In three out of four objects, accuracy was lower for degenerate views learned in a spatiotemporally predictable context compared to a completely unpredictable one.

2.4.2 A behavioral disadvantage for spatial prediction during object learning

Intuitively, spatial predictability should be advantageous for learning the three-dimensional structure of objects given that it is the learning context within which the visual system evolved. However, the literature contains a mixture of contradictory effects regarding the utility of spatially predictable information during object learning and recognition. Initial experiments described in Lawson et al. (1994) with depth-rotated line drawings of familiar objects demonstrated the expected increase in recognition accuracy for spatially predictable sequences. A number of studies have found that studying depth-rotating sequences of novel objects under one ordering and then testing with a different ordering impairs recognition (Stone, 1998; Vuong & Tarr, 2004; Chuang et al., 2012), implying that learned predictability about the sequence is used to encode the object (Balas & Sinha, 2009a). The foreshortening model advanced in (Balas & Sinha, 2009b) also provided a better match to observers' data by incorporating spatial information (e.g., the first- and second-order derivatives of the foreshortening function over object views).

Some of the experiments described in Lawson et al. (1994), however, demonstrated better accuracy for sequences studied with weak spatial predictability (maximum of two spatially coherent frames in the sequence) than total spatial predictability. Experiments described in Harman and Humphrey (1999) failed to find any positive or negative effects of spatial predictability on accuracy. They did, however, increase in reaction time for objects learned in a spatially predictable

context, similar to the one reported here. One possible reason for the behavioral disadvantage for objects learned with spatial predictability is that less attention is necessary in these conditions. A constantly changing sequence of views might require more attentive processing to encode whereas the relatively low amount of change between views in spatially predictable sequences is comparatively "unsurprising" such that some views might be overlooked during encoding. However, there was some indication that the adverse effect of spatial predictability was driven primarily by the degenerate views of the stimuli used in the present work. A more focused experiment is clearly necessary to explicitly test the hypothesis that behavioral performance is impaired for degenerate views learned in a spatially predictable context but relatively intact for canonical views.

2.4.3 Viewpoint invariance for "paper clip" objects

The "paper clip" objects used in the current work have a long history of use in studies of three-dimensional viewpoint effects in human observers (Bulthoff & Edelman, 1992; Edelman & Bulthoff, 1992; Sinha & Poggio, 1996) as well as studies monkey physiology studies (Logothetis et al., 1994; Logothetis et al., 1995). The objects are easy to generate systematically and thus can be combined with a staircase procedure to titrate difficulty or can be generated en masse to find the parameters that elicit maximal responses from neurons. Various effects with the objects have been replicated using computational models of object recognition with identical stimuli (Riesenhuber & Poggio, 1999) and mathematical properties of the objects are known to capture a large amount of variability in behavior (Balas & Sinha, 2009b). Thus, it can be reasonably concluded that paper clip objects are a useful class of stimuli.

However, other work brings under question the ecological validity of paper clip objects. The objects are constructed from thin line segments separated by empty space and thus self-occlusion of features is less of a problem than for three-dimensional volumetric objects with surfaces. This might imply that view invariance is not actually necessary to represent the full three-dimensional structure of paper clip objects, since the majority of features can be extracted from a static view. Accordingly, studies comparing three-dimensional objects composed of line segment with volu-

metric objects found that the line segment objects were not represented in a viewpoint invariant manner (Farah, Rochlin, & Klein, 1994; Pizlo & Stevenson, 1999).

Thus, it is is possible that a spatiotemporally predictable training context is simply not optimal for learning to represent paper clip objects. Temporal association mechanisms (Stringer et al., 2006; Wallis & Baddeley, 1997; Isik et al., 2012) might bias the development of viewpoint invariance by forming unwanted associations between canonical and degenerate views, which could lower overall accuracy levels and slow reaction times. This explanation would require that these mechanisms were somehow not elicited in the unpredictable learning contexts.

Chapter 3

Neural model of spatiotemporal prediction for object recognition

3.1 Introduction

The work presented in this chapter describes a neural network model of the broader LeabraTI framework (Chapter **??**). The specific implementation was used to investigate the role of spatiotemporal predictive learning in an object recognition task, analogous to the ones used in the Chapter 1 and 2 experiments. The principal behavioral results of these experiments are first reviewed before turning to the model implementation and simulations that reproduce these results.

The Chapter 1 experiment investigated the role of predictive processing during a novel object recognition task. The experiment made use of novel three-dimensional "paper clip" objects that required integration over multiple sequential views to extract their three-dimensional structure. The results of the experiment indicated that spatial and temporal predictability of an entraining sequence enhanced discriminability of a subsequently presented probe stimulus using a same-different judgement.

The Chapter 2 experiment expanded on the previous chapter's experiment by investigating the role that spatial and temporal predictability played during prolonged learning about the same paper clip objects. The experiment involved an explicit training period during which observers studied the objects while they were rotated through their views followed by a series of test trials that required same-different judgements about static probe stimuli. Somewhat surprisingly, the results of the experiment were an almost complete reversal of the previous chapter's experiment. Accuracy was lowest when stimuli were learned in a combined spatially and temporally predictable

context and highest when learned in a completely unpredictable context.

TODO: Foreshadow results...

3.2 Methods

3.2.1 Model architecture

The model architecture is illustrated in Figure 3.1. The model consisted of three layers and one preprocessing stage whose parameters are described in detail in the following paragraphs. Two of the layers contained columnar substructure necessary for learning using the LeabraTI algorithm . To simplify the overall LeabraTI computation, only superficial (Layers 2 and 3) and deep (Layer 6) neuron subtypes were explicitly modeled. Projections between these neuron populations corresponded to the descending Layer 5 $\rightarrow$ Layer 6 synapses in the brain, which are assumed to be plastic, and the ascending Layer 6 $\rightarrow$ Layer 4 transthalamic synapses which are assumed to be relatively stable and nonplastic.

Retina and V1 preprocessing: Input was provided to the model via a 24x24 topographic filter bank that preprocessed images offline from the model proper. This preprocessing step is consistent with a large class of biological models describing object recognition in cortex (e.g., Riesenhuber & Poggio, 1999; Serre et al., 2007; O'Reilly et al., 2013) and in the case of the present model, represents visual processing from the level of the retina through V1 simple cells (Hubel & Wiesel, 1962). Grayscale bitmap images were scaled to 24x24 pixels and convolved with Gabor filters at four orientations (0°, 45°, 90°, and 135°) and two polarities (off-on and on-off) producing a 24x24x4x2 set of inputs. Each Gabor filter was implemented as 6x6 pixel kernel, with a wavelength λ = 6 and Gaussian width terms of σ_x = 1.8 and σ_y = 1.2. A static nonlinearity was applied to the output of the filtering step in the form of a modified *k*-Winners-Take-All (*k*WTA) inhibitory competition function that reduced activation across the 4x2 filter bank to the equivalent of k = 1 fully active units (see O'Reilly et al., 2013, Supporting Information).

Primary visual layers: 24x24 topographic layer arranged into groups of 4x2 units (4608

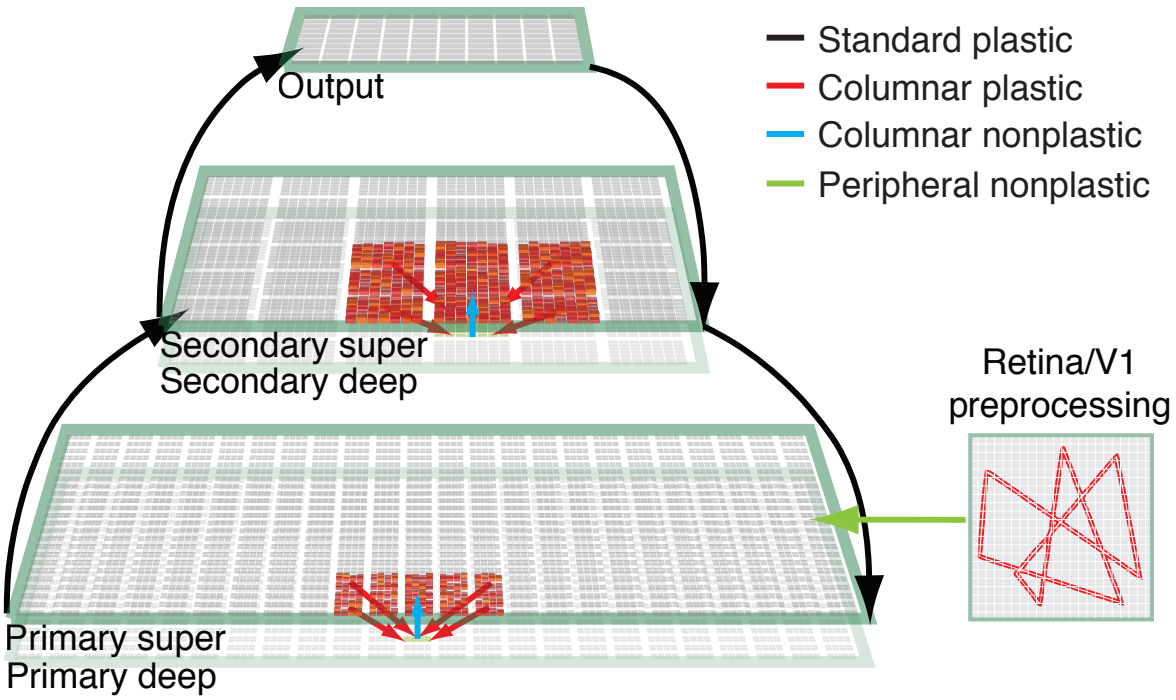

Figure 3.1: Model architecture

The model's four layers and principal projections. Primary and secondary visual layers contained columnar substructure in which deep units integrated from 5x5 columns of superficial units in the primary case or 3x3 columns in the secondary case. Ascending synapses from deep to superficial units were nonplastic and connected in a one-to-one manner.

total units), decomposed into superficial and deep neuron subtypes. Each superficial unit received the output of the retina/V1 preprocessing step. kWTA inhibition for superficial units was set to 60% of the average of the top k active units compared to the average of all other superficial units with each 4x2 unit group using a value of $k = 2$. Deep units received from 5x5 columns of superficial units (200 inputs per deep unit) integrated into a single value that was used as the additional context input channel for each superficial unit.

Secondary visual layers: 6x6 topographic layer arranged into groups of 7x7 units (1764 total units), also decomposed into superficial and deep neuron subtypes. Each superficial unit received from 8x8 topographical neighborhoods of early visual columns (512 afferents per unit) and sent back reciprocal connections with the same topography. kWTA for superficial units was

set to 60% of the average of the top k active units compared to the average of all other superficial units with each and 15% activity within each unit group. Deep units received from 3x3 columns of superficial units (441 inputs per deep unit) integrated into a single value that was used as the additional context input channel for each superficial unit.

Output layer: 10x10 layer (100 total units) without unit group or columnar substructure. Each unit received a full projection from secondary visual columns (1764 afferents per unit) and fully projected back to all columns. A scale of 10% was used to limit the influence of the output units on secondary visual columns during the training period, preventing "hallucinatory" representations that can become disconnected from bottom-up inputs. A kWTA value of $k = 1$ was used to enforce a localist representation. The localist representation is a computational simplification that allowed an identity readout of lower-level features without population decoding similar to that provided by inferior temporal (IT) neurons (Hung, Kreiman, Poggio, & DiCarlo, 2005; Li, Cox, Zoccolan, & DiCarlo, 2009).

3.2.2 LeabraTI learning algorithm

LeabraTI was implemented as an extension of the standard Leabra algorithm which is described in detail in O'Reilly and Munakata (2000) and O'Reilly, Munakata, Frank, Hazy, and Contributors (2012). Standard Leabra learning operates across two phases: a *minus* phase that represents the system's expectation for a given input and a *plus* phase, representing observation of the outcome. The difference between the minus and plus phases, along with additional self-organizing mechanisms, is used in computing the synaptic weight update function at the end of each plus phase.

LeabraTI extends standard Leabra learning by interleaving its minus and plus phases over temporally contiguous input sequences. In standard Leabra, the minus phase depends on clamped inputs from the sensory periphery to drive the expectation while the plus phase uses clamped outputs from other neural systems to drive the outcome. In LeabraTI, the minus phase expectation is not driven by the sensory periphery, but instead by lagged context represented by deep (Layer

6) neurons. During the plus phase, driving potential shifts back to the sensory periphery. Deep neurons' context is also updated after each plus phase.

LeabraTI was only used to update the synaptic weights between superficial and deep neurons. Inter-areal feedforward and feedback projections bifurcate from the local column, directly synapsing disparate populations of superficial neurons and thus weight updates in these cases were handled by standard Leabra equations. In computing the weight update, the standard Leabra delta rule (O'Reilly, 1996) uses the difference in rate between the plus and minus phases of receiving units (y) in proportion to the rate of sending units (x) in the minus phase:

$$\Delta_{leabra} w_{ij} = x^{-}(y^{+} - y^{-})$$

In the LeabraTI framework, deep neurons are considered to be the receiving units as they are the terminus of the descending columnar synapses. However, deep units are proposed to only be active during the minus phase when they drive the prediction, and thus cannot be used to compute an error signal. To address this issue, we invert the LeabraTI delta rule:

$$\begin{aligned} \Delta_{leabrati} w_{ij} &= super^{-}(deep^{+} - deep^{-}) \\ &\approx deep^{-}(super^{+} - super^{-}) \end{aligned}$$

Additionally, the temporally extended nature of the algorithm requires that the receiving units represent the current state (time t) and sending units the previous moment's state (time t - 1). While conceptualized as the previous equation, the actual implementation is as follows:

$$\Delta_{leabrati} w_{ij} = super^{+}_{t-1}(super^{+}_{t} - super^{-}_{t})$$

This formulation allows the driving potential of deep neurons to be computed just once using the previous plus phase state of superficial neurons (multiplied by the superficial $\rightarrow$ deep learned weights) and held constant as an input to superficial neurons during the minus phase. This is a gross simplification of the actual biological process of deep neurons, but is vastly more computationally efficient than explicit modeling by computing an additional rate for each deep neuron at each

time step. This formulation is also equivalent to the simple recurrent network (SRN) (Elman, 1990; Servan-Schreiber, Cleeremans, & McClelland, 1991), thus providing a potential biological substrate for its computational function.

One limitation of LeabraTI's interleaving of minus and plus phases over time is that the initial minus phase in an input sequence does not have access to the previous moment's context. Even if there was lagged context available, it would represent information from a prior, possibly unrelated input sequence. To address this, weight updates are disabled for the first minus-plus phase pair, and enabled thereafter. In the brain, this process might be facilitated by a neural mechanism that is sensitive to the repetition of inputs over time (e.g., acetylcholine) (Thiel, Henson, Morris, Friston, & Dolan, 2001; Thiel, Henson, & Dolan, 2002).

3.2.3 Training and testing environment

LeabraTI requires training to establish establish the spatial associations over subsequent time steps. In human development, this is expected to be facilitated by coarse transformations of retinal inputs due to environmental or self motion. This initial learning stage develops generic features that capture how inputs change from moment-to-moment (100 ms timeframes in LeabraTI). The actual inputs are not critical except that they accurately reflect the average statistics of the environment. In training the model, a simplified "paper clip" environment was assumed, using the four objects from the Chapter 2 experiment.

During training, an input sequence depicted one of the four objects rotating coherently through all 30 view renderings (adjacent views spaced 12 degrees apart). During the minus phase, the model made a prediction about the upcoming view and during the plus phase, the view was processed by the retina/V1 filter banks and clamped as an input to the model. The output unit corresponding to each object was also clamped during the plus phase to bias views belonging to the same object toward similar lower-level feature representations. Training proceeded for 20 epochs of 10 randomly selected input sequences each. The learning rate on all plastic synapses started at 1.0 and was halved every 8 epochs.

Training efficacy was evaluated by computing the average cosine (normalized dot product) between the minus and plus phase for the primary and secondary visual layers:

$$cos\theta = \frac{1}{n}\sum_{k=1}^{n}\frac{layer_k^- \cdot layer_k^+}{||layer_k^-||||layer_k^+||}$$

The cosine varies between 0 and 1 and expresses the degree of similarity of LeabraTI's prediction to the actual outcome in layers with columnar substructure (Figure 3.2). A value of zero indicates the minus phase prediction is completely orthogonal to the plus phase sensation and a value of 1 indicates complete overlap. The lower-bound on the cosine is not likely to be zero as it would require spurious activations in topographic regions that do not contain any features. A better approximation of the lower bound is the case when the system simply reproduces the plus phase from the previous moment's (*t* - 1) state. This can be thought of as the amount of perceptual overlap between adjacent views of the stimuli, and thus any additional features that contribute to a higher cosine value indicate positive prediction.

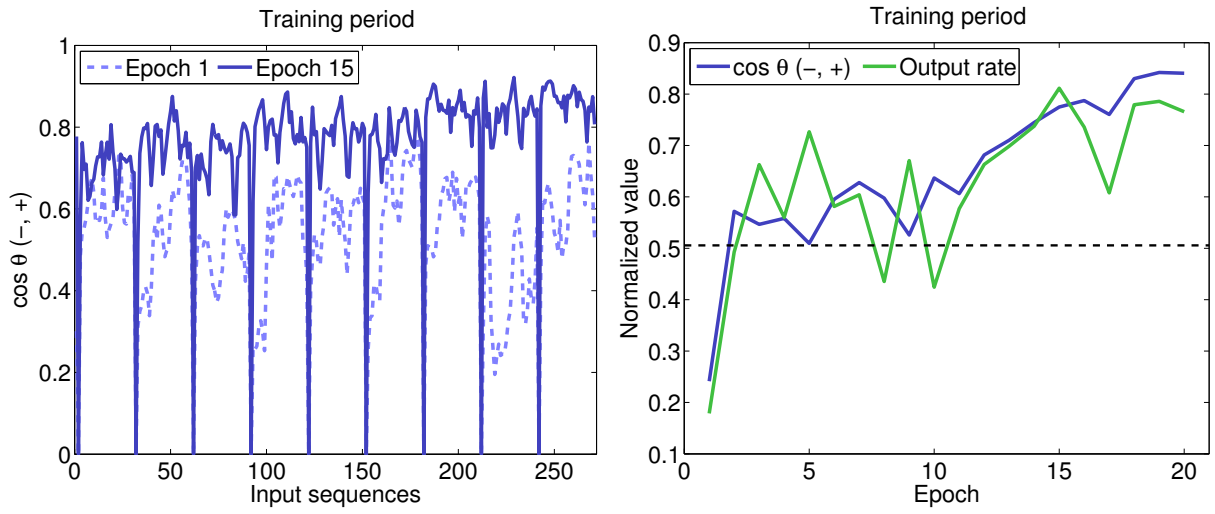

Figure 3.2: Model training

Average cosine between minus and plus phase for layers with columnar substructure and output response rate over the course of training. Sharp drops in the cosine to zero indicate the start of a new input sequence and are unlearnable. The lower bound for the cosine was computed as the reproduction of the plus phase from the previous moment's (*t* - 1) state and the overall average is indicated by the dotted line. A cosine greater than this level indicates positive prediction.

Typically, after the initial feature training phase, neural models are trained to classify stimulus-response pairs (Riesenhuber & Poggio, 1999; Serre et al., 2007; although, see also O'Reilly et al., 2013). In human learning, stimulus predictability and response mappings can be learned independently (Wyart et al., 2012; Kok et al., 2012). The present model was compact and input environment simple enough that the initial features and response mappings could be learned jointly. The rate of the target output unit was used to evaluate the efficacy of the learned response mappings.

Consistency between features and responses was ensured by using two sets of synapses with different update intervals that contribute a weighted mixture to the input of each receiving unit. The first "standard" set of synapses were updated after every plus phase, whereas a second "stable" set of synapses were updated at the end of each epoch. In the present model, a 80% stable to 20% standard synaptic mixture was used. This allowed the model to more slowly integrate learning across an entire epoch's worth of input sequences without runaway representations caused by being exposed to the same rotating object's features over multiple time steps while still maintaing the moment-to-moment spatiotemporal predictive learning central to LeabraTI.

Testing involved presenting input sequences accordant with each of the four predictability conditions used in the Chapter 1 and 2 experiments. In the spatially unpredictable conditions (S-), random views were selected for each plus phase and used to compute deep neurons' updated context. To model the effect of temporal unpredictability (T-), a variable number of time steps (up to four) separated each context update. Each time the context update was skipped, a decay factor of 50% was applied to superficial neurons' context input channel. The default scale of this channel was 100% and thus four time steps without a context update decayed the scale to 12.5%. The net effect of temporal unpredictability was a weakening of the prediction at each time step until the next view was actually presented and the updated context could be computed.

The completely unpredictable condition (S-T-) utilized both the variable update interval and decay whereas the combined spatial and temporal predictability condition (S+T+) was identical to the training procedure (i.e., a coherently rotating object with constant update interval). In all

cases, predictions about each upcoming view were made during each minus phase given the current context state. Weight updates that normally occurred at the end of each plus phase during training were disabled on all plastic synapses during testing.

3.3 Results and Discussion

The results from the Chapter 1 and 2 experiments along with the results of the model test sequences are plotted in Figure 3.3. *d'* (sensitivity) was used as the common behavioral measure across experiments due to the issues with response bias in raw accuracy found in the Chapter 1 experiment. The rate of the target output unit was used to compare the model with the experimental results. All model results reflect the weights learned after 15 training epochs, as this was the point when the output rate was maximal, allowing for the largest potential differences between conditions during testing. This epoch choice also mitigated overfitting issues given the relatively simple training environment.

The Chapter 1 experiment tested subjects' ability to differentiate objects that were presented after a short series of spatiotemporally predictable or unpredictable entraining views. Subjects only ever saw 168 degrees of an object spread across 8 views on any given trial. Although feedback was given following the response on each trial, the relatively short exposure to disparate object views combined with the relatively large set of 16 possible target objects likely discouraged substantial learning. The trained model without modifications was capable of producing these results. Output unit rate was super-additive in the combined spatial and temporal predictability case as the testing sequence perfectly matched the training environment in terms of spatial and temporal properties and thus maximally activated both superficial and deep units. This is a fundamental prediction of the LeabraTI model that was not present in the experimental results, but has been reported previous investigations of predictability on attentional allocation (Doherty et al., 2005; Rohenkohl et al., 2014).

The Chapter 2 experiment produced an almost complete reversal of the results from the Chapter 1 experiment. This second experiment differed from the first a number of meaningful

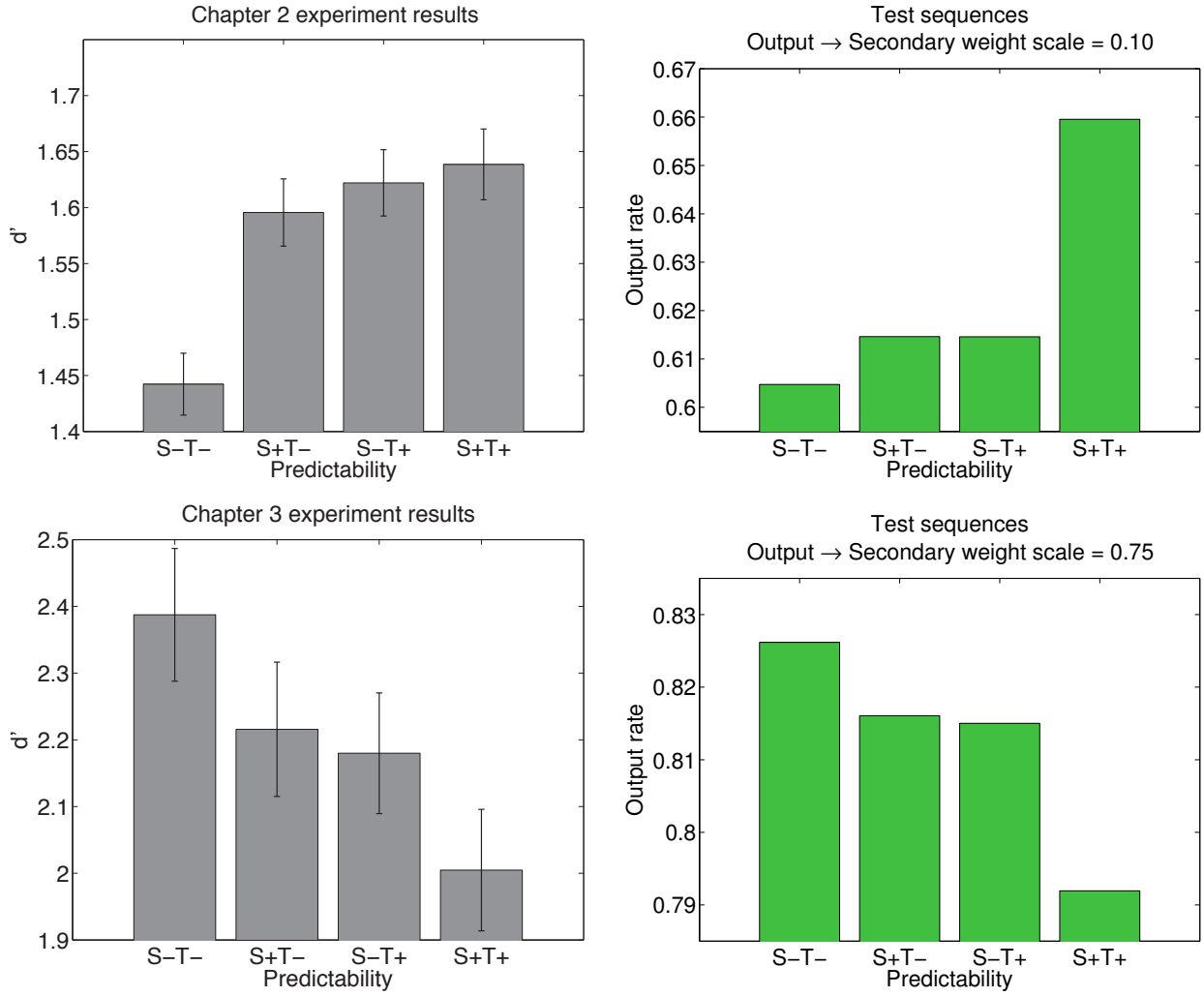

Figure 3.3: Experiment and modeling results

Chapter 1 (**top**) and 2 (**bottom**) experiment and model results. *d'* (sensitivity) was used as the common behavioral measure across experiments and the rate of the models' target output unit was used in comparison.

ways. First, a smaller set of only four target objects was used. Subjects also observed each of the objects rotate completely through each of its views four times and were explicitly instructed to study the object as it rotated. No feedback was given during test trials, but each object was seen four separate times during the experiment and many subjects reported being aware of the fact that there were four unique objects. A reasonable conclusion is that these differences encouraged over-training of the objects and that spatial and temporal predictability interact with this overtraining in

different ways.

LeabraTI is predicated on spatiotemporal regularity and is thus somewhat inappropriate for evaluating learning under spatially and temporally unpredictable contexts. To account for the Chapter 2 results, a simple proxy was used for overtraining the stimuli in which the scale of the weights on the Output $\rightarrow$ Secondary visual synapses was increased. Typically, a relative scale of 10% is used on feedback projections so that feedforward inputs drive the majority of weight changes with feedback playing a more modulatory role (Crick & Koch, 1998; Sherman & Guillery, 1998). This is crucial for the training period to prevent "hallucinatory" representations that can become disconnected from bottom-up inputs and produces the best testing results since model adapts its weights to the strength of inputs for each layer.

Increasing the scale of the weights on the Output $\rightarrow$ Secondary visual synapses to 75% produced the same reversal observed in the Chapter **??** results in which training in the combined spatial and temporal predictability context impaired recognition relative to the completely unpredictable case. Synaptic weight scaling is one of the many effects of learning, especially when considering the long timescale self-organizing mechanisms presumed by Leabra that reinforce the most active units (O'Reilly & Munakata, 2000; O'Reilly et al., 2012). The full range of the reversal effect when increasing Output $\rightarrow$ Secondary visual synaptic weight scale is plotted in Figure 3.4A. Overall, the effect is graded and thus varying the amount of exposure observers have with to stimuli would probably modulate prominence of the reversal effect.

To determine the effect of learning on the representation of the objects, the cosine was used to compute a pairwise similarity metric over secondary visual unit minus phase activations across all views of all objects (i.e., representational similarity, Kriegeskorte et al., 2008). LeabraTI training produced a representation that captures some similarity across sequential views but each view remained relatively distinct, as would be expected of V2-level representations (Kobatake & Tanaka, 1994; Freeman & Simoncelli, 2011). The proxy for learning used here strengthens the synapses between secondary visual units and higher-level areas that code increasingly invariant representation. In the model, this higher-level area was a localist output layer which can be considered to be

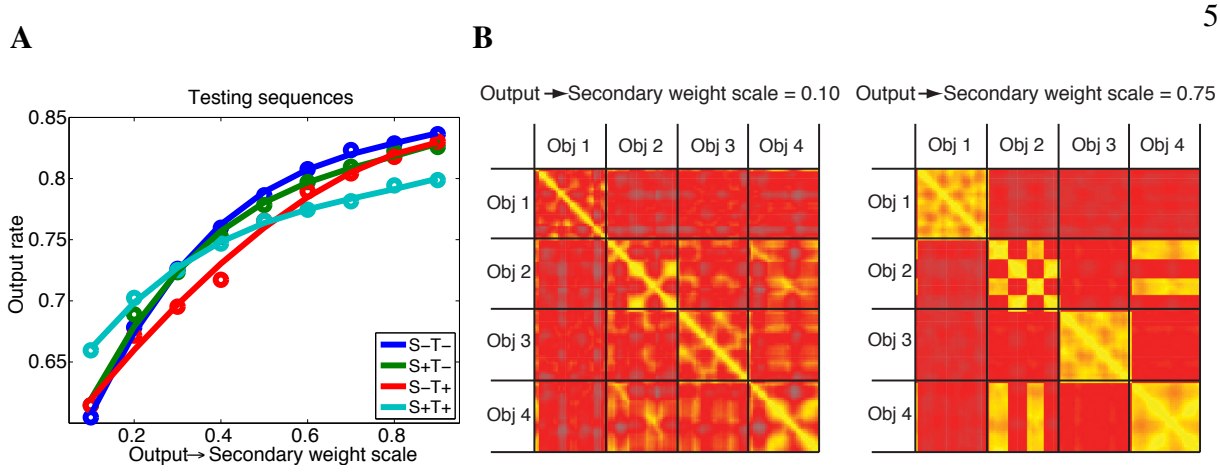

Figure 3.4: Effect of prolonged learning and representational similarity

A: Target output rate as a function of Output → Secondary visual synaptic weight scale. Lines indicate best fit third order polynomials. **B:** Pairwise cosine over secondary visual unit minus phase activations across all views of all objects. Yellow indicates greater similarity. Results shown for both 10% and 75% Output → Secondary visual weight scales.

coding the same invariant representation that IT cortex does using a population code (Hung et al., 2005; Li et al., 2009).

The representational similarity suggests that prolonged learning causes invariance to "trickle down" to lower-level feature representations. This is problematic for objects that suffer from severely degenerate views such as Object 2.[1] For Object 2, two distinct views were represented, divided by the degenerate view. However, one of these views was represented similarly to Object 4. This object confusion was less of an issue when the objects were recently acquired (10% Output → Secondary visual weight scale) and might account for the comparatively lower performance of objects studied for prolonged periods with spatiotemporal predictability.

[1] In the Chapter 2 experiment, accuracy for Object 2 suffered the most of all objects for degenerate views, falling from ceiling to below chance levels.

References

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. Trends in Cognitive Sciences, 16(7), 390–398.

Balas, B., & Sinha, P. (2009a). Learned prediction affects body perception. Visual Cognition, 17(5), 679–699.

Balas, B. J., & Sinha, P. (2009b). The role of sequence order in determining view canonicality for novel wire-frame objects. Attention, Perception & Psychophysics, 71(4), 712–723.

Belyusar, D., Snyder, A. C., Frey, H.-P., Harwood, M. R., Wallman, J., & Foxe, J. J. (2013). Oscillatory alpha-band suppression mechanisms during the rapid attentional shifts required to perform an anti-saccade task. NeuroImage, 65, 395–407.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics, 29(4), 1165–1188.

Brainard, D. (1997). The Psychophysics Toolbox. Spatial Vision, 10(4), 433–436.

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proceedings of the National Academy of Sciences of the United States of America, 89(1), 60–64.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. The Journal of Neuroscience, 29(24), 7869–7876.

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. Proceedings of the National Academy of Sciences of the United States of America, 107(37), 16048–16053.

Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (in press). Entrainment of neural oscillations as a modifiable substrate of attention. Trends in Cognitive Sciences.

Capotosto, P., Babiloni, C., Romani, G. L., & Corbetta, M. (2009). Frontoparietal cortex controls spatial attention through modulation of anticipatory alpha rhythms. The Journal of Neuroscience, 29(18), 5863–5872.

Chuang, L. L., Vuong, Q. C., & Bulthoff, H. H. (2012). Learned non-rigid object motion is a view-invariant cue to recognizing novel objects. Frontiers in Computational Neuroscience, 6(26).

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Massons method. Tutorials in Quantitative Methods for Psychology, 1(1), 42–45.

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. Nature Neuroscience, 8(9), 1145–1147.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. The Journal of Neuroscience, 33(9), 4002–4010.

Crick, F., & Koch, C. (1998). Constraints on cortical and thalamic projections: The no-strong-loops hypothesis. Nature, 391(6664), 245–250.

Dayan, P., Hinton, G. E., Neal, R. N., & Zemel, R. S. (1995). The Helmholtz machine. Neural Computation, 7(5), 889–904.

de Graaf, T. A., Gross, J., Paterson, G., Rusch, T., Sack, A. T., & Thut, G. (2013). Alpha-band rhythms in visual task performance: Phase-locking by rhythmic sensory stimulation. PLoS One, 8(3).

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9–21.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? Neuron, 73(3), 415–434.

Dien, J. (2009). A tale of two recognition systems: Implications of the fusiform face area and the visual word form area for lateralized object recognition models. Neuropsychologia, 47(1), 1–16.

Doherty, J. R., Rao, A., Mesulam, M. M., & Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. The Journal of Neuroscience, 25(36), 8259–8266.

Edelman, S., & Bulthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vision Research, 32(12), 2385–2400.

Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14(2), 179–211.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. Journal of Cognitive Neuroscience, 19(9), 1488–1497.

Farah, M., Rochlin, R., & Klein, K. (1994). Orientation invariance and geometric primitives in shape recognition. Cognitive Science, 18(2), 325–344.

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., & Molholm, S. (2011). Ready, set, reset: Stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. The Journal of Neuroscience, 31(27), 9971–9981.

Foldiak, P. (1991). Learning invariance from transformation sequences. Neural Computation, 3(2), 194–200.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. Experimental Brain Research, 142(1), 139–150.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. Nature Neuroscience, 14(9), 1195–1201.

Friston, K. (2005). A theory of cortical responses. Philosophical Transactions of the Royal Society B, 360(1456), 815–836.

George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. PLoS Computational Biology, 5(10).

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. Nature Neuroscience, 15(4), 511–517.

Gould, I. C., Rushworth, M. F., & Nobre, A. C. (2011). Indexing the graded allocation of visuospatial attention using anticipatory alpha oscillations. Journal of Neurophysiology, 105(3), 1318–1326.

Harman, K. L., & Humphrey, G. K. (1999). Encoding 'regular' and 'random' sequences of views of novel three-dimensional objects. Neuropsychologia, 28(5), 601–615.

Horschig, J. M., Jensen, O., van Schouwenburg, M. R., Cools, R., & Bonnefond, M. (2013). Alpha activity reflects individual abilities to adapt to the environment. NeuroImage, 89, 235–243.

Hubel, D., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. Journal of Physiology, 160(1), 106–154.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science, 310(5749), 863–866.

Isik, L., Leibo, J. Z., & Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. Frontiers in Computational Neuroscience, 6(37).

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. Vision Research, 49(10), 1295–1306.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway. Journal of Neurophysiology, 71(3), 856–867.

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. Cerebral Cortex, 22(9), 2197–2206.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron, 60(6), 1126–1141.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. Human Brain Mapping, 8(4), 194–208.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. Current Biology, 22(11), 1000–1004.

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions: Evidence for viewpoint-specific recognition procedures. Perception, 23(5), 595–614.

Lee, T. S., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. Journal of the Optical Society of America, 20(7), 1434–1448.

Li, N., Cox, D., Zoccolan, D., & DiCarlo, J. (2009). What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? Journal of Neurophysiology, 102(1), 360–376.

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. Science, 321(5895), 1502–1507.

Li, N., & DiCarlo, J. J. (2010). Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. Neuron, 67(6), 1062–1075.

Li, N., & Dicarlo, J. J. (2012). Neuronal learning of invariant object representation in the ventral visual stream is not dependent on reward. The Journal of Neuroscience, 32(19), 6611–20.

Logothetis, N., Pauls, J., Bulthoff, H., & Poggio, T. (1994). View-dependent object recognition by monkeys. Current Biology, 4(5), 401–414.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Current Biology, 5(5), 552–563.

Mathewson, K., Gratton, G., Fabiani, M., Beck, D., & Ro, T. (2009). To see or not to see: Prestimulus alpha phase predicts visual awareness. The Journal of Neuroscience, 29(9), 2725–2732.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. Cognition, 115(1), 186–191.

Mathewson, K. E., Prudhomme, C., Fabiani, M., Beck, D. M., Lleras, A., & Gratton, G. (2012). Making waves in the stream of consciousness: Entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation. Journal of Cognitive Neuroscience, 24(12), 2321–2333.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. Proceedings of the National Academy of Sciences of the United States of America, 108(48), 19401–19406.

Nowak, L., & Bullier, J. (1997). The timing of information transfer in the visual system. In K. S. Rockland, J. H. Kaas, & A. Peters (Eds.), Cerebral Cortex: Volume 12. Extrastriate Cortex in Primates (pp. 205–241). New York, New York: Plenum Press.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience, 2011.

O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. Neural Computation, 8(5), 895–938.

O'Reilly, R. C., & Munakata, Y. (2000). Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. Cambridge, MA: The MIT Press.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors (2012). Computational Cognitive Neuroscience. Wiki Book, 1st Edition, URL: `http://ccnbook.colorado.edu`.

O'Reilly, R. C., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. J. (2013). Recurrent processing during object recognition. Frontiers in Psychology, 4(124).

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10(4), 437–442.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. Electroencephalography and Clinical Neurophysiology, 72(2), 184–187.

Pizlo, Z., & Stevenson, A. K. (1999). Shape constancy from novel views. Perception & Psychophysics, 61(7), 1299–1307.

Posner, M. I. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32(1), 3–25.

Rao, R. P., & 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.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11), 1019–1025.

Rohenkohl, G., Gould, I. C., Pessoa, J., & Nobre, A. C. (2014). Combining spatial and temporal expectations to improve visual perception. Journal of Vision, 14(4), 1–13.

Rohenkohl, G., & Nobre, A. C. (2011). Alpha oscillations related to anticipatory attention follow temporal expectations. The Journal of Neuroscience, 31(40), 14076–14084.

Romei, V., Gross, J., & Thut, G. (2012). Sounds reset rhythms of visual cortex and corresponding human visual perception. Current Biology, 22(9), 807–813.

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. Nature, 354(6349), 152–155.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. Trends in Cognitive Sciences, 12(3), 106–113.

Schultz, J., Brockhaus, M., Bulthoff, H. H., & Pilz, K. S. (2013). What the human brain likes about facial motion. Cerebral Cortex, 23(5), 1167–1178.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proceedings of the National Academy of Sciences of the United States of America, 104(15), 6424–6429.

Servan-Schreiber, D., Cleeremans, A., & McClelland, J. L. (1991). Graded state machines: The representation of temporal contingencies in simple recurrent networks. Machine Learning, 7(2–3), 161–193.

Sherman, S., & Guillery, R. (1998). On the actions that one nerve cell can have on another: Distinguishing "drivers" from "modulators". Proceedings of the National Academy of Sciences of the United States of America, 95(12), 7121–7126.

Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. Nature, 384(6608), 460–463.

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. Vision research, 38(7), 947–951.

Stringer, S. M., Perry, G., Rolls, E. T., & Proske, J. H. (2006). Learning invariant object recognition in the visual system with continuous transformations. Biological Cybernetics, 94(2), 128–142.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nature Neuroscience, 11(9), 1004–1006.

Thiel, C. M., Henson, R. N., Morris, J. S., Friston, K. J., & Dolan, R. J. (2001). Pharmacological modulation of behavioral and neuronal correlates of repetition priming. The Journal of Neuroscience, 21(17), 6846–6852.

Thiel, C. M., Henson, R. N. A., & Dolan, R. J. (2002). Scopolamine but not lorazepam modulates face repetition priming: A psychopharmacological fMRI study. Neuropsychopharmacology, 27(2), 282–292.

Townshend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. N. Castellan Jr., & F. Restle (Eds.), Cognitive Theory: Volume 3 (pp. 200–239). Hillsdale, NJ: Lawrence Erlbaum Associates.

Townshend, J. T., & Ashby, F. G. (1983). Stochastic Modeling of Elementary Psychological Processes. Cambridge: Cambridge University Press.

VanRullen, R., & Dubois, J. (2011). The psychophysics of brain rhythms. Frontiers in Psychology, 2(203).

Vuong, Q. C., & Tarr, M. J. (2004). Rotation direction affects object recognition. Vision Research, 44(14), 1717–1730.

Wallis, G., Backus, B. T., Langer, M., Huebner, G., & Bulthoff, H. (2009). Learning illumination- and orientation-invariant representations of objects through temporal association. Journal of Vision, 9(7).

Wallis, G., & Baddeley, R. (1997). Optimal, unsupervised learning in invariant object recognition. Neural Computation, 9(4), 883–894.

Wallis, G., & Bulthoff, H. (1999). Learning to recognize objects. Trends in Cognitive Sciences, 3(1), 22–31.

Wallis, G., & Bulthoff, H. (2001). Effects of temporal association on recognition memory. Proceedings of the National Academy of Sciences of the United States of America, 98(8), 4800–4804.

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. Proceedings of the National Academy of Sciences of the United States of America, 109(9), 3593–3598.