Chapter 1

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

1.1 Introduction

TODO: No real new motivation – just say something about spatial sequences and how they are highly specific per object. Ref Balas's work, primarily

1.2 Methods

1.2.1 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.

1.2.2 Stimuli

Novel "paper clip" objects were used as stimuli (see Chapter **??** 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 **??** experiment. Target and distractor objects were paired together for the purposes of the experiment. All objects are shown in Figure 1.1.

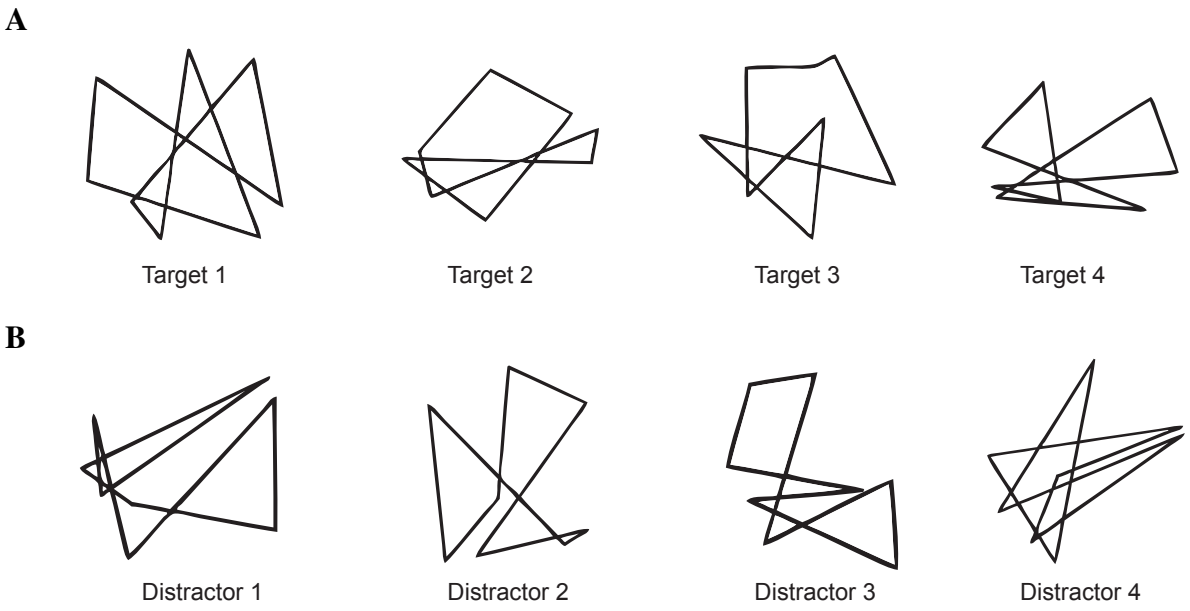

Figure 1.1: Novel "paper clip" objects

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

1.2.3 Procedure

The experiment was divided into 16 blocks, each containing a training period followed by a series of test trials (Figure 1.2). 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).

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 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.

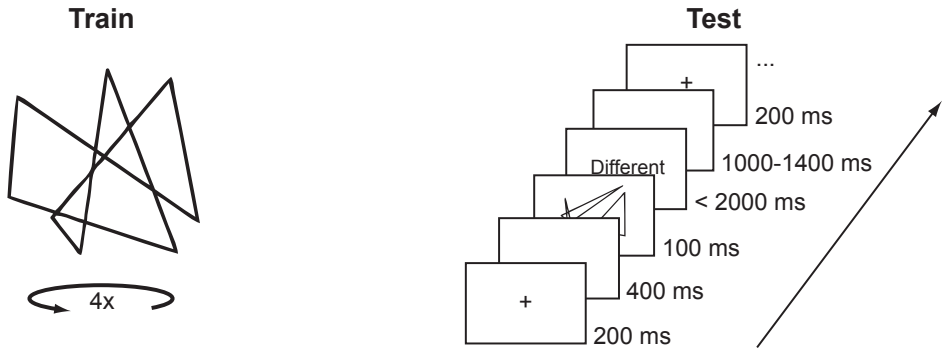

Figure 1.2: Experimental procedure

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.

1.3 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 are plotted in Figure 1.3. Transformed behavioral measures (e.g., *d'*, inverse efficiency; see Chapter **??** Results) showed similar patterns to the raw measures and were thus omitted from analysis.

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 be-

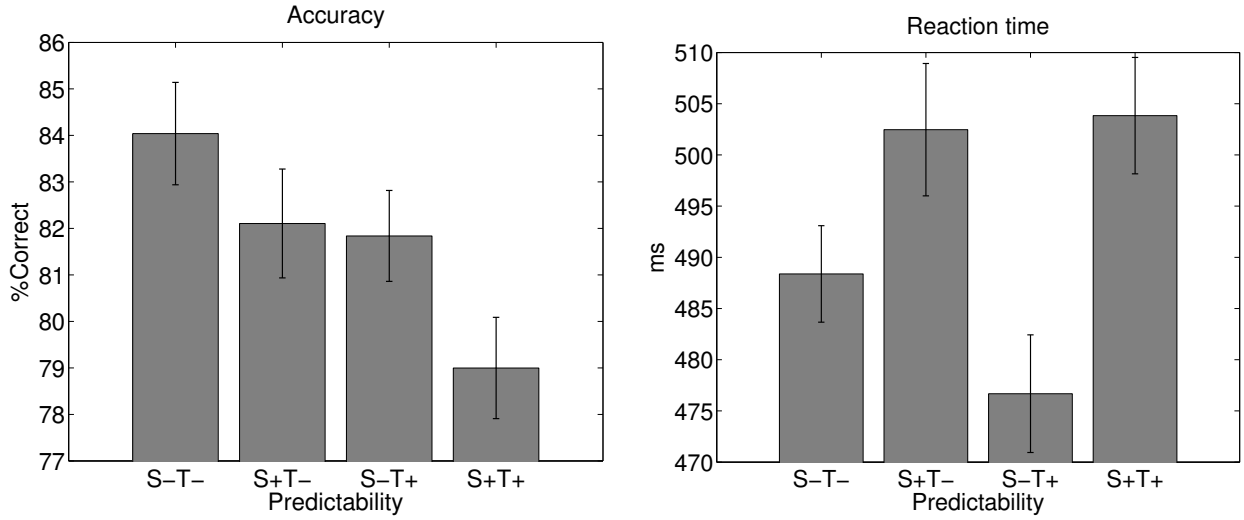

Figure 1.3: Behavioral measures of spatial and temporal predictability

Accuracy and reaction time 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.

tween 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 1.3). 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+, $t(58) = -1.77$, $p = 0.082$).

Subjects were also slower to respond when the training period was spatially predictable ($F(1, 57) = 10.99$, $p = 0.002$). 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 were highly variable across target objects (Figure 1.4). Target-condition assignment did not significantly affect accuracy or reaction times (both 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, 2009). 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.

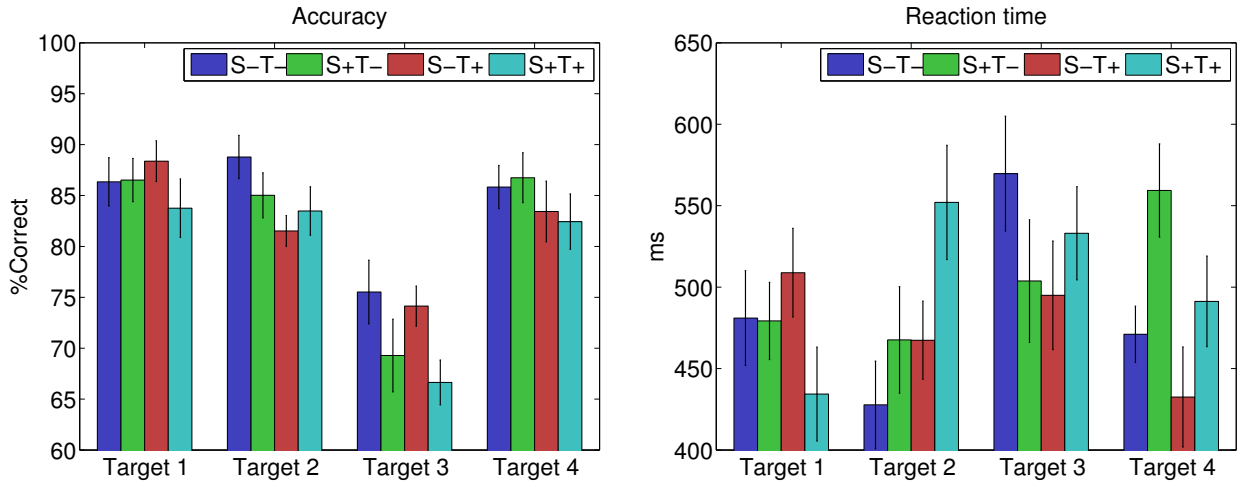

Figure 1.4: Behavioral measures for each target object

Accuracy and reaction times for each target object. Horizontal axes denote target and colors predictability during the training period. Error bars depict between-subjects standard error.

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 1.5). Test trials during which distractor objects were presented were 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

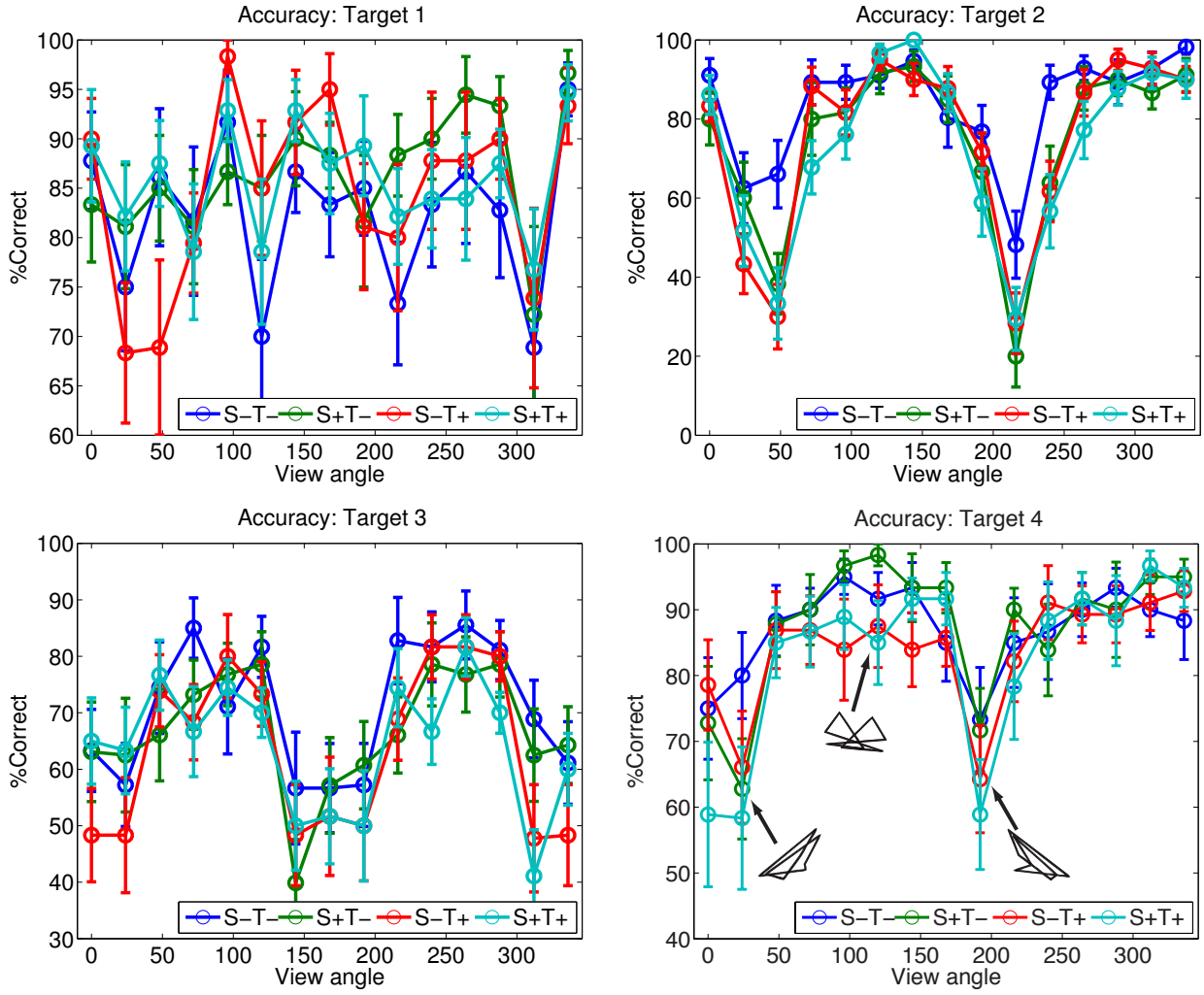

Figure 1.5: Accuracy as a function of viewing angle for each target object

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.

for degenerate views because each data point corresponded to only four trials per subject. To address this design limitation, a bootstrapping method was used to resample the available data in these cases. The accuracy function over viewing angles was collapsed across conditions and the two minima associated with degenerate views were identified for each object. For target object 1, this corresponded to angles $\theta = \{24°, 312°\}$, object 2: $\theta = \{48°, 240°\}$ object 3: $\theta = \{144°, 312°\}$,

and object 4 $\theta = \{24^\circ, 192^\circ\}$. Accuracy for the completely unpredictable (S-T-) and combined spatial and temporal predictability (S+T+) conditions during training was averaged at these viewing angles and resampled with replacement from the 59 subjects for 10000 iterations. This produced distributions for degenerate view accuracy for each object (Figure 1.6). Accuracy for was lower for degenerate views for the combined spatial and temporal predictability condition for all target objects except target 1, which didn't exhibit the patterned accuracy function that other targets did. The S+T+/S-T- 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.

1.4 Discussion

1.4.1 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 predictability. 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 differences between predictability conditions during training were driven primarily by degenerate viewing angles caused by three-dimensional foreshortening in the objects used. A foreshortening model has previously accounted for the principal component of variability in recognition accuracy for the same "paper clip" objects used here (Balas & Sinha, 2009), but only in a combined spatial and temporal predictability learning context.

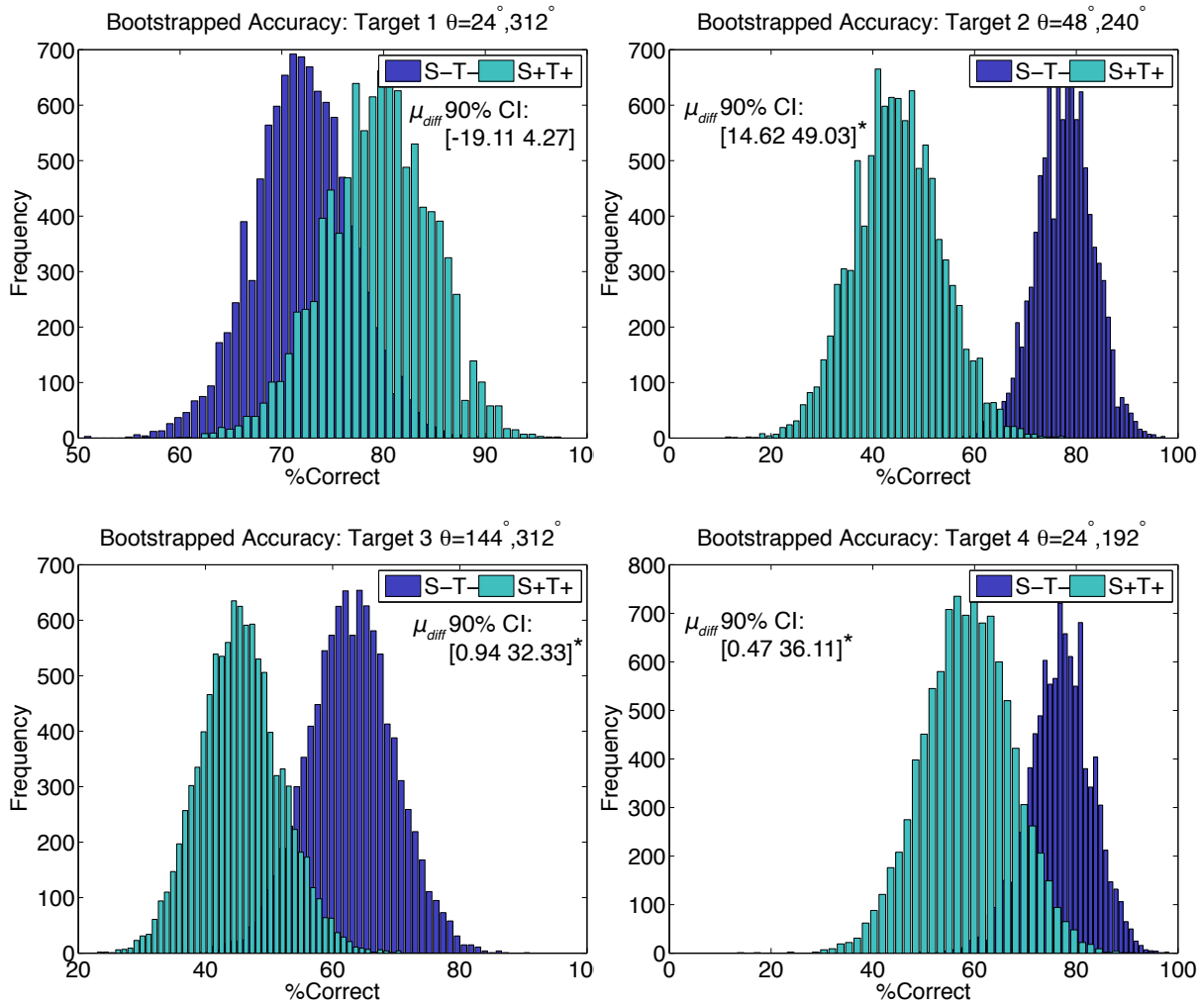

Figure 1.6: Bootstrapped accuracy for degenerate views

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.

1.4.2 A behavioral disadvantage for spatial prediction during object learning

Based on previous research, it is unclear whether spatial predictability is used by the brain during object learning and whether it is actually advantageous. Initial investigations described in Lawson, Humphreys, and Watson (1994) identified the expected increase in recognition accuracy for spatially predictable sequences. The foreshortening model advanced in (Balas & Sinha, 2009) was also improved by incorporating spatial information (e.g., the first- and second-order deriva-

tives of the foreshortening function over object views). Furthermore, a number of computational models, including LeabraTI (Chapter **??**), use learning rules that incorporate associations between subsequent inputs to learn object representations (Foldiak, 1991; Stringer, Perry, Rolls, & Proske, 2006; Wallis & Baddeley, 1997; Isik, Leibo, & Poggio, 2012).

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 slowing of reaction times 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 less "surprising" and 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 would be necessary to explicitly test the hypothesis impaired for degenerate views learned in a spatially predictable context and relatively intact for canonical views, but the interpretation of this hypothesis if confirmed is still discussed here.

1.4.3 Spatiotemporal prediction biases development of invariance

The theory advanced here is that spatially predictable sequences promote the development of invariance over the sequence transformation given prolonged learning. For example, if a three-dimensional object rotates in depth in a spatially predictable manner, associations can be formed between subsequent views using a temporal association rule. Integrating over small changes in viewing angle is easier than large changes (Logothetis, Pauls, Bulthoff, & Poggio, 1994; Logothetis, Pauls, & Poggio, 1995) and thus, the problem of constructing invariance can be solved gradually instead of all at once. A large body of previous work supports this idea.

Behavioral experiments by Wallis, Bulthoff, and colleagues have used a predictability paradigm for studying face recognition similar to the one used in the present work, but in which spatially

unpredictable sequences are characterized by swapping the identity of faces mid-sequence. Most observers were unaware of these identity swaps, but they significantly impaired the discriminability of swapped identities compared to stable identities (Wallis & Bulthoff, 2001). The effects were originally reported for identities swapped during depth-rotated sequences but have since been extended to swaps during changes in orientation and illumination (Wallis, Backus, Langer, Huebner, & Bulthoff, 2009). Together, these results suggest that associations are made between subsequent members of a sequence to construct invariance to transformations.

Single unit recordings from Li and DiCarlo using a similar swap paradigm have shown exactly how this invariance is constructed. Li and DiCarlo (2008) Li and DiCarlo (2010)

References

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

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

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.

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

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

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).

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.

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.

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.

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

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.

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.

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

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.