

Neural Correlates of Implicit Knowledge about Statistical Regularities

Sachio Otsuka<sup>1,2</sup>, and Jun Saiki<sup>1</sup>

<sup>1</sup>*Japan Society for the Promotion of Science*

<sup>2</sup>*Kyoto University*

Author Note

Sachio Otsuka, Japan Society for the Promotion of Science, Tokyo, Japan; Graduate School of Human and Environmental Studies, Kyoto University, Japan; now at Faculty of Culture and Information Science, Doshisha University, Japan; Jun Saiki, Graduate School of Human and Environmental Studies, Kyoto University, Japan.

Correspondence concerning this article should be addressed to Sachio Otsuka, Faculty of Culture and Information Science, Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe-shi, Kyoto-fu 610-0394, Japan (e-mail: [sotsuka@mail.doshisha.ac.jp](mailto:sotsuka@mail.doshisha.ac.jp)).

### Abstract

In this study, we examined the neural correlates of implicit knowledge about statistical regularities of temporal order and item chunks using functional magnetic resonance imaging (fMRI). In a familiarization scan, participants viewed a stream of scenes consisting of structured (i.e., three scenes were always presented in the same order) and random triplets. In the subsequent test scan, participants were required to detect a target scene. Test sequences included both forward order of scenes presented during the familiarization scan and backward order of scenes (i.e., reverse order of forward scenes). Behavioral results showed a learning effect of temporal order in the forward condition and scene chunks in the backward condition. fMRI data from the familiarization scan showed the difference of activations between the structured and random blocks in the left posterior cingulate cortex including the retrosplenial cortex. More important, in the test scan, we observed brain activities in the left parietal lobe when participants detected target scenes on temporal order information. In contrast, the left precuneus activated when participants detected target scenes based on scene chunks. Our findings help clarify the brain mechanisms of implicit knowledge about acquired regularities.

*Key words:* statistical learning, implicit knowledge, posterior cingulate cortex, parietal lobe, precuneus.

## Neural Correlates of Implicit Knowledge about Statistical Regularities

**Introduction**

It is important for us to detect targets in natural environments rapidly and accurately. We may feel we can easily detect targets in our daily life. Consistent with this intuition, many previous studies have reported ultra-rapid natural scene categorization and scene perception in the near absence of selective attention (e.g., Li et al. 2002; Otsuka and Kawaguchi 2007; Thorpe et al. 1996). Recently, it has been demonstrated that scene perception is supported not only by associative learning of contextual information (e.g., Davenport and Potter 2004) but also by prior exposure to statistical regularities (i.e., statistical learning). Statistical learning is defined as the extraction of systematic regularities embedded in a continuous stream of auditory or visual stimuli (e.g., Fiser and Aslin 2001, 2002; Saffran et al. 1996). A typical visual statistical learning (VSL) experiment consists of familiarization and test phases. In a familiarization phase, participants observe a visual stream consisting of shapes (e.g., Fiser and Aslin 2002; Turk-Browne et al. 2005, 2008; Turk-Brown and Scholl 2009), letters (Turk-Browne et al. 2009), objects (e.g., Otsuka et al. 2013, 2014), or natural scenes (e.g., Brady and Oliva 2008; Otsuka et al. 2016). The visual stream consists of temporal triplets that three items always appear in the same order across participants (e.g., ABC, DEF, GHI, and JKL; “A” is always followed by “B”, which is always followed by “C”). In the next test phase, participants perform a two-alternative forced-choice (2AFC) familiarity decision. On each test trial, participants observe two, three-item test sequences presented sequentially. One test sequence is a triplet of items that have been presented during the familiarization phase (e.g., ABC). Another test sequence is a foil constructed with items from three different triplets (e.g., AEI). Participants are required to decide whether the first or second test sequence seems more familiar based on the familiarization stream. A number of previous studies have shown

that participants correctly choose the triplets above chance (i.e., 50%), suggesting that people extract and learn statistical regularities of temporal order information.

Recently, it has been focused on whether the expression of VSL reflects implicit or explicit knowledge about statistical information. Aside from imposing explicit familiarity judgments, Kim et al. (2009) used a rapid serial visual presentation (RSVP) target detection task as an implicit measure. Following a typical familiarization phase, participants performed an RSVP target detection task where they were first asked to remember a target shape, and then detect the target shape as quickly as possible in a continuous stream of random presentations of four triplets. The rationale was that if participants learned the statistical regularities of the triplets during the familiarization phase, reaction times (RTs) for target detection should be shorter for the second and third predicted triplet shapes than for the first predictive shape. Kim et al. showed that RTs for the shapes presented in the third triplet position were shorter than those in the first triplet position, even when they performed an explicit matching task where one isolated shape was presented, and they could not identify the two shapes that formed a triplet among the 11 shapes displayed during the familiarization phase.

More recently, with the familiarity test and target detection task as direct and indirect measures of VSL, respectively, Batterink et al. (2015) have shown that auditory statistical learning produces both implicit and explicit knowledge about acquired statistical regularities. Following this previous study, Otsuka et al. (2016) examined whether VSL produced implicit and/or explicit knowledge about temporal order information and scene chunks (e.g., aggregative scene representations which do not have temporal order information). To examine whether the expression of VSL in the target detection was based on temporal order of scenes or scene chunks, Otsuka et al. (2016) used the same (forward) order of scenes as presented during the familiarization phase and reverse (backward) order of scenes (e.g.,

CBA; see Turk-Browne and Scholl, 2009). Evidence of VSL in the forward condition was assumed to reflect learning primarily based on serial order information acquired during the familiarization phase. In contrast, the VSL occurring in the backward condition was expected to reflect effects of aggregative scene information. That is, Otsuka et al. (2016) predicted that if participants acquired implicit knowledge about chunks, RTs would be faster for targets in positions 2 and 3 compared to position 1 in the backward condition. In contrast, RTs would not differ between positions 2 and 3 given that the predictive power of “C” in position 1 and “B” in position 2 would be similar in the backward condition. In other words, “C” in position 1 could predict either “A” or “B” in position 2 based on chunk representation, and “B” in position 2 could predict either “A” or “C” in position 3, which would result in similar RTs. Otsuka et al. showed a learning effect for both temporal order in the forward condition and chunks in the backward condition of the target detection task. Their results suggested that VSL produced implicit knowledge about learned temporal order and scene chunks.

The purpose of the current study is to clarify brain mechanisms underlying implicit knowledge about temporal order information and item chunks in VSL. Some previous studies have identified the brain areas engaged in VSL. Using functional magnetic resonance imaging (fMRI), Turk-Browne et al. (2010) examined the neural relationship between the predictive (“A” elements in “AB” pairs) and predicted items (“B” elements) in VSL. They used an online VSL paradigm in which participants discriminated face-sex (male or female faces) pairs, or scene categories (indoor or outdoor scenes) while observing a stream of scene-face or face-scene pairs. Results indicated that the anterior hippocampus showed greater responses to predictive versus unpaired stimuli, indicating anticipation as a consequence of VSL. In addition, hippocampal anticipation based on predictive images correlated with subsequent processing of the predicted images in the left inferior parietal lobule (IPL). Thus, we can predict that implicit knowledge about temporal order information

may manifest as brain activity changes in the anterior hippocampus and IPL. In contrast, to our knowledge, few studies have focused on the brain mechanisms underlying implicit knowledge about item chunks. Although learning in this study was not strictly chunk learning, Schapiro et al. (2013) demonstrated that people could learn to create event clusters or a community structure of mutually predicting visual events. This occurs when participants are told to parse a stream of meaningless visual letters or objects using subjective judgments of breaking points in sequences without uniform transitional probabilities. Importantly, brain activities in the left inferior frontal gyrus (IFG), insula, anterior temporal lobe, and superior temporal gyrus, engaged in learning a community structure, were found in participants who did not actually parse visual stimuli into a cluster. Thus, it is possible that implicit knowledge about item chunks may elicit activation in these brain areas. In the current study, we concurrently explored neural correlates of implicit knowledge about temporal order information and item chunks in a single experiment.

Also, we assessed brain activity when participants learn statistical regularities embedded in a visual stream of a familiarization scan. Comparing the structured blocks where three letters always appeared in the same order with random blocks without consistent regularities, Turk-Browne et al. (2009) measured brain activity during the familiarization, and observed activation in the right hippocampus and the right caudate. In addition, we examined whether the parahippocampal place area (PPA), specific to processing of natural scenes, is engaged in implicit knowledge about temporal order information and scene chunks. It has been demonstrated that three brain regions are related to processing images of familiar locations: the PPA, retrosplenial cortex, and transverse occipital sulcus (e.g., Epstein et al., 2007). Recently, Turk-Browne et al. (2012) demonstrated that neural activity in the PPA is modulated by temporal context (e.g., a photo of a forest was consistently followed by an image of a bathroom). In the current study, similar to previous VSL studies using natural

scenes (Turk-Browne et al., 2010; Turk-Browne et al., 2012), we used the PPA as our region-of-interest (ROI).

## **Material and methods**

### **Participants**

Twenty-four participants from the participant pool of Kyoto University took part in this experiment (13 males and 11 females, mean age: 22.2 years; age range: 20-27 years). They participated in this experiment in exchange for a book coupon (5,000 yen). Data from all participants were used in the analyses. All participants were right-handed based on H. N. handedness inventory (Hatta, 1975), and they had self-reported normal or corrected-to-normal visual acuity with non-magnetic glasses. Finally, all had normal color vision based on the Ishihara test for color blindness. All experimental protocols were approved by the Institutional Review Board of Kyoto University. All participants signed informed consent forms.

### **Stimuli**

Twenty-four natural scenes were selected from the SUN database (Brady and Oliva, 2008; Xiao et al., 2010; <http://groups.csail.mit.edu/vision/SUN/>). They belonged to 24 different categories (bar, bathroom, bedroom, bridge, building, cave, cemetery, coast, conference room, desert, field, forest, gas station, golf course, house, kitchen, living room, mountain, restaurant, street, swimming pool, tennis court, waterfall, and wave). These images were presented in color on a black background. Half of the images were used for four structured triplets in the structured blocks, whereas the remaining images were used for four random triplets in the random blocks. We created 12 stimulus lists to counterbalance across participants. In addition, for a localizer scan, we selected 12 additional natural scenes (six indoor scenes and six outdoor scenes) from the same database as the main scan, and 12 faces (six male faces and six female faces) from Matsumoto and Ekman (1988). Visual stimuli were displayed on a projection screen at the back of the scanner bore: they were viewed with



a mirror attached to the head coil. The experiment was controlled by the time schedule of Presentation software (Neurobehavioral Systems).

### **Procedure**

Following Turk-Browne et al. (2009), there were 24 (12 structured and 12 random blocks) 18-sec blocks in the familiarization scan, which included a rest period and a 800-ms digit display with a 200-ms ISI. Each block contained four structured, or four random triplets, presented in random order. In the structured blocks, three natural scenes consistently appeared in the same order within each triplet (e.g., ABC, DEF, GHI, JKL). Across participants, each natural scene from 12 images was assigned to a unique position within one of the four structured triplets. In contrast, in the random blocks, the remaining 12 natural scenes were randomly presented in four random triplets with the following constraint: to maintain the same positional structure as structured triplets, each image always appeared in the same position within any of four random triplets. To wit, to control positional encoding, each scene appeared only in position 1 (e.g., A'-D'-G'-J'), position 2 (e.g., B'-E'-H'-K'), or position 3 (e.g., C'-F'-I'-L') of the four random triplets. Block order was counterbalanced across participants. Each of the twelve natural scenes in the block was presented individually for 800 ms followed by a 200-ms ISI. Then, a 5-sec rest period was inserted. In addition, one display containing three numbers (e.g., 364) was inserted in a random position after any one of the four triplets in each block, for 800 ms followed by a 200-ms ISI (see Fig. 1). During this scan, participants were required to detect the number display and press the button as quickly as possible. Participants were never informed that there was any structure in the stream of structured blocks.

Following the familiarization scan, participants took part in the test scan consisting of four runs. On each trial, a target scene was presented for 2,000 ms, and participants were required to remember it. Following a 1,000 ms blank screen, five natural scenes were

presented individually for 200 ms, with the ISI of 800 ms (see Fig. 2). Of these five images, the second, third, and fourth images were from the structured blocks in the familiarization scan, and they consisted of triplets or foils. The first and fifth images were from the random blocks in the familiarization scan (i.e., filler images). Across four runs, each target scene from the structured blocks was presented 12 times, in the second, third, or fourth position of triplets and foils. Each image from the random blocks was presented 8 times, in the first or fifth position. The rest period was 4 sec. Participants were required to press the button with the index finger of their right hand as quickly as possible when they detected the target natural scene.

In each test run, each target scene was presented three times in the second, third, or fourth position of triplets and foils, and 12 targets were presented in the fifth position (filler trials), resulting in 48 trials. Target scenes were never presented in the first position. The order of target position for each target scene was randomized across three repetitions. Each run contained four types of blocks: (1) forward triplets with three natural scenes that appeared in the same order as those during the familiarization scan (ABC, DEF, GHI, and JKL); (2) forward foils used in most VSL studies (AEI, DHL, GKC, and JBF); (3) backward triplets (CBA, FED, IHG, and LKJ) and (4) backward foils (IEA, LHD, CKG, and FBJ) where the original triplets or foils were presented in reverse order. Thus, participants completed 192 trials in total through four runs. The order of blocks within each run was counterbalanced across participants.

After the test scan, participants took part in one run of a localizer scan. Each block consisted of 12 faces or scenes, of natural scenes that were not used in the main experiment. The two categories alternated every 24 sec. The order of blocks was counterbalanced across participants. Participants judged whether faces were male or female, or whether scenes occurred indoors or outdoors by pressing one of the two buttons. On each block, a fixation

cross was displayed for 6 sec with an instruction about finger/key constraints. Then, each image was presented for 200 ms with a 1,300-ms ISI, resulting in 24 sec blocks. There were eight blocks of each type, resulting in 16 blocks.

After the localizer scan, participants were asked: “Did you notice any patterns in the visual stream of natural scenes in the first scan?” None of participants reported noticing any statistical regularity.

### **fMRI data acquisition**

Data were collected on a 3T Siemens scanner (3.0T MAGNETOM Verio) at Kokoro Research Center, Kyoto University. Functional data were acquired with a T2\* weighted gradient-echo, EPI sequence (TE = 25 msec; TR = 2,000 msec; flip angle = 75°; matrix = 64 × 64; field of view = 224 mm; 3.5 mm thickness) with 34 axial slices. We acquired 216 volumes for the familiarization run that lasted 432 sec, 288 volumes for each test run that lasted 576 sec each, and 192 volumes for the localizer run that lasted 384 sec. Structural images were acquired using a T1-weighted anatomical sequence (MPRAGE).

### **Preprocessing**

Preprocessing and statistical analyses were performed using Statistical Parametric Mapping software (SPM; <http://www.fil.ion.ucl.ac.uk>) in MATLAB (The MathWorks, Inc., Natick, MA). Images were corrected for head motion with realignment to the first volume, corrected for slice acquisition time, spatially normalized using the Montreal Neurological Institute (MNI) EPI template, and spatially smoothed using an 8-mm Gaussian kernel.

### **Data analyses**

To examine how brain activities differed with respect to the forward-triplet and backward-triplet blocks in the test scan, we ran a general linear model (GLM) analysis using SPM. Each block type (i.e., forward triplet, forward foil, backward triplet, and backward foil) was entered as a separate regressor. As covariates of no interest, we included six regressors

for each dimension of head motion. Each regressor was convolved with the canonical hemodynamic response function. As the period of each trial lasted 8 sec, we set the duration of each regressor for four scans. These models estimated the contribution of each block type to the BOLD response in every voxel for each participant. In addition, to examine how brain activities differed between the structured and random blocks in the familiarization scan, we ran an additional GLM analysis, which focused on learning of statistical regularities in the familiarization scan (the duration of each regressor was seven scans). The resulting parameter estimates (beta values) for the two conditions in each scan were compared across participants using paired  $t$  tests. A cluster of voxels was judged to show a reliable difference for the contrasts of forward versus backward blocks and triplet versus foil blocks in the test scan, and a reliable difference for the contrast of structured versus random blocks in the familiarization scan, if the associated  $t$  value reached significance at the cluster level threshold  $p < 0.05$  corrected (the voxel level threshold  $p < 0.001$  uncorrected). We labeled each region using the Talairach Client (Lancaster et al., 1997, 2000) after converting from MNI to Talairach coordinates (<http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach>).

In addition, to examine whether statistical structure modulated brain activity in a scene specific region, we compared BOLD responses for forward versus backward blocks in the test scan, and for structured versus random blocks in the familiarization scan within one *a-priori* ROI from the localizer scan: bilateral PPA. Natural scene images were contrasted against face images to localize the PPA. In each region, the voxel with the greatest  $t$  value in an anatomically restricted search was selected as the center of an 8-mm sphere ROI. We used the MarsBaR ROI toolbox to calculate percent signal changes (Brett et al., 2002). Percent signal changes were calculated separately in the left and right PPA, and these data were collapsed. Finally, to examine the association between behavioral data from target detection

and brain activation, we conducted a whole brain analysis using the difference in RTs between triplet and foil conditions in each position as a covariate.

## Results

### Behavioral Results

First, we analyzed the behavioral data. In the familiarization scan, participants detected the number display with few errors (99.7%,  $SE = 0.2$ ), and close to no false alarms. These results demonstrate that participants adequately observed the visual stream of natural scenes. Next, we analyzed RT data of the test scan. We excluded from the following analyses trials with misses and errors (in this case, pressing the button before a target image was presented), and also other trials as outliers with RTs less than 100 milliseconds (ms), greater than 1,000 ms, and deviated by at least 2  $SDs$  from the mean RT. This trimming procedure was independently performed for each participant and condition. This procedure resulted in excluding 3.0% of the trials as misses and errors, and 6.5% of the trials as RT outliers.

Fig. 3 shows RTs in the target detection task during the test scan. We conducted a three-way within-subjects ANOVA involving the triplet temporal order (forward vs. backward), type of sequences (triplet vs. foil), and target position (position 1, position 2, vs. position 3). We failed to find a significant two-way interaction,  $F(2, 46) = 1.17, p = 0.319, \eta^2_p = 0.049$ . Given that some previous studies have revealed learning effects only for the triplet sequences in the target detection task (e.g., Bertels et al., 2012; Otsuka et al., 2016), we performed a series of planned  $t$ -test comparisons (the Bonferroni correction). These analyses revealed that RTs for positions 2 (528 ms) and 3 (506 ms) were shorter than those for position 1 (555 ms) in the forward-triplet condition,  $t(23) = 3.25, p = 0.011$  and  $t(23) = 5.26, p < 0.001$ , respectively; moreover, RTs for position 3 were shorter than those for position 2,  $t(23) = 2.96, p = 0.021$ . In the backward-triplet condition, RTs for positions 2 (516 ms) and 3 (521 ms) were shorter than those for position 1 (558 ms),  $t(23) = 5.28, p < 0.001$  and  $t(23) = 3.67, p = 0.004$ , respectively. However, in the backward condition, the difference in RTs between positions 2 and 3 was not significant,  $t(23) = 0.95, p = 1.000$ . More importantly, we

observed a significant difference (i.e., 22 ms vs. -5 ms) between the difference in RTs between position 2 and position 3 of the forward-triplet condition (22 ms) and that in RTs between position 2 and position 3 of the backward-triplet condition (-5 ms),  $t(23) = 3.99$ ,  $p = 0.001$ . On the other hand, we did not observe learning effects either in the forward-foil (position 1: 547 ms; position 2: 550 ms; position 3: 541 ms; position 1 vs. position 2:  $t(23) = 0.47$ ,  $p = 1.000$ ; position 1 vs. position 3:  $t(23) = 0.61$ ,  $p = 1.000$ ; position 2 vs. position 3:  $t(23) = 1.18$ ,  $p = 0.748$ ) or backward-foil conditions (position 1: 564 ms; position 2: 555 ms; position 3: 554 ms; position 1 vs. position 2:  $t(23) = 1.67$ ,  $p = 0.326$ ; position 1 vs. position 3:  $t(23) = 1.50$ ,  $p = 0.444$ ; position 2 vs. position 3:  $t(23) = 0.10$ ,  $p = 1.000$ ).

Taken together, the pattern of ANOVA results is consistent with the findings reported by Otsuka et al. (2016). The present data suggest that VSL produced implicit knowledge about the statistical regularities of temporal order information and scene chunks.

## **fMRI results**

### **Whole-Brain Analysis**

#### *Statistical learning in the familiarization scan*

To explore the neural correlates of statistical learning, we conducted a voxel-wise contrast of structured versus random blocks in the familiarization scan. Only the left posterior cingulate cortex (peak of mass in MNI coordinates, -12, -56, 16;  $p = 0.041$  corrected; extent = 264; peak  $t = 5.18$ ) responded strongly to random versus structured triplets (see Fig. 4 and Table 1). Note that this area also contains the retrosplenial cortex (-12, -55, 11), which is more specific to scene perception (e.g., Epstein et al., 2007; Park & Chun, 2009). In fact, the retrosplenial cortex in the whole brain analysis overlapped with the scene selective region reported by Epstein et al. (2007).

#### *Implicit knowledge about learned temporal order information and scene chunks*

More importantly, we conducted a voxel-wise contrast of forward versus backward blocks in the test scan, to examine the neural correlates of implicit knowledge about temporal order information. The left parietal lobe ( $-30, -44, 24$ ;  $p < 0.001$  corrected; extent = 367; peak  $t = 4.85$ ) responded more strongly to the target detection on the predictable forward ordered scenes than the backward ordered scenes (i.e., forward triplet > backward triplet; see Fig. 5 and Table 1). In addition, we conducted a voxel-wise contrast of triplet versus foil blocks in the test scan, to examine the neural correlates of implicit knowledge about scene chunks. As a result, the left precuneus ( $-22, -58, 50$ ;  $p = 0.011$  corrected; extent = 287; peak  $t = 4.46$ ) responded more strongly to target detection from the triplet sequences, including information of scene chunks, than to that from the foil sequences, not including any statistical information (i.e., forward triplet + backward triplet > forward foil + backward foil; see Fig. 6 and Table 1).

### **ROI Analysis**

According to Turk-Browne et al. (2010), we examined the effect of implicit knowledge about statistical structure in bilateral PPA ROIs, identified in each participant from the functional localizer. Pertinent ROIs were found bilaterally in all participants (mean MNI coordinates, right: 24,  $-42, 8$ ; left:  $-28, -58, -18$ ). We conducted a two-way within-subjects ANOVA involving triplet temporal order and type of sequences on the data of the test scan (see Table 2), and observed the significant main effect of type of sequence,  $F(1, 23) = 7.24, p = 0.013, \eta^2_p = 0.239$ . Mean percent signal changes in the triplet condition (0.798) were lower than those in the foil condition (0.828). However, neither the effect of type of sequences nor interaction were significant,  $F < 1$ , and  $F(1, 23) = 1.11, p = 0.304, \eta^2_p = 0.046$ , respectively. No significant difference was observed in signal change between structured and random blocks in the familiarization scan,  $t < 1$ .<sup>1</sup>

### **Relationship between brain activations and behavioral performance**



To examine the relationship between behavioral data of target detection in the test scan and brain activations during the familiarization and test scans, we conducted the same whole brain analyses as previously described (i.e., on the fMRI data from familiarization and test scans), using either the difference in RTs (35 ms in the forward condition, and 33 ms in the backward condition, respectively) between position 3 of the forward-foil (541 ms) and forward-triplet conditions (506 ms) or that between position 3 of the backward-foil (554 ms) and backward-triplet conditions (521 ms) as a covariate. These differences in RTs would reflect learning effects in each order condition. This analysis revealed that the left precuneus region ( $-12, -56, 58$ ;  $p < 0.001$  corrected; extent = 662; peak  $t = 6.07$ ), and right caudate ( $14, 20, 14$ ;  $p = 0.036$  corrected; extent = 255; peak  $t = 5.69$ ) were activated greater for structured versus random blocks in the familiarization scan, with the difference in RTs between position 3 of the forward foil and triplet conditions as a covariate. In contrast, we did not observe any correlation between behavioral data and neural activation in the test scan.

## **Discussion**

### **Target detection in the test scan**

First, RTs for natural scenes presented in positions 2 and 3 were shorter than those presented in position 1 in the forward-triplet condition. In addition, RTs for scenes in position 3 were shorter than RTs for scenes in position 2. This pattern of results is consistent with the findings reported by the previous studies (Bertels et al., 2012; Kim et al., 2009; Otsuka et al., 2016; Turk-Browne et al., 2005). In contrast, in the backward-triplet condition, although RTs for scenes in position 2 and 3 were shorter than those in position 1, there was no difference in RTs between positions 2 and 3. In short, behavioral results are consistent with Otsuka et al. (2016) in that VSL produced implicit knowledge about learned statistical regularities of temporal order information and scene chunks. It should be noted that because we did not ask the participants to do a familiarity decision as a direct measure of VSL, we could not examine whether a learning effect on the target detection task was found not only for the remembered triplets but also unremembered triplets. However, considering some previous studies (Batterink et al., 2015; Otsuka et al., 2016), it is reasonable to argue that a learning effect found in the target detection task reflect implicit knowledge about acquired statistical regularities of temporal order and scene chunks.

### **Brain activities to implicit knowledge about temporal order information and scene chunks**

The left parietal lobe, including the IPL and supramarginal gyrus in this cluster, was strongly activated when participants detected the target scenes from the forward sequence of natural scenes (i.e., the contrast between forward-triplet and backward-triplet blocks in target detection). These results converge with those of Turk-Browne et al. (2010) wherein participants completed the ongoing classification task on learned temporal order information. Furthermore, Karuza et al. (2013) observed activation in the left supramarginal gyrus when

participants listened to forward versus backward speech. The left supramarginal gyrus reported in our study ( $-35, -49, 28$ ) may overlap with that reported by Karuza et al. (2013) given the relatively large size of clusters in their study ( $-52, -48, 10$ ; extent = 4,085). In other words, our findings suggest that the left parietal lobe, including the IPL and supramarginal gyrus, is engaged in implicit knowledge about learned temporal order information. In addition, Maffei et al. (2015) showed that watching the walking movie in the backward order activated the bilateral intraparietal sulci (IPS) and suggested that activity in the IPS reflects the difficulty arising with unfamiliar stimuli such as impossible walking (i.e., visual walking in the backward order). Therefore, ours and the previous findings suggest that the parietal region may be differentially involved in implicit knowledge about temporal order information and may be sensitive to unfamiliar sequences based on prior experience.

In contrast, we observed significant activity in the left precuneus when participants detected the targets from the triplet sequences, including information of scene chunks (i.e., the contrast between both forward- and backward-triplet blocks and both forward- and backward-foil blocks). Recently, it has been reported that the precuneus was more strongly activated in professional than in amateur players when they perceived board patterns during a board game named Shogi (Wan et al., 2011). Wan et al. (2011) suggested that the precuneus is engaged in perception of spatial patterns such as chunk information. In addition, the ROI analysis showed a difference of signal changes in the PPA between the triplet and foil blocks, which is comparable to the findings of previous studies that showed signal changes in the PPA for items bearing statistical information (e.g., Turk-Browne et al., 2010, 2012). Our results and the previous findings suggest that the precuneus and PPA may be engaged in perception of item chunks, including statistical regularities.

Taken together, we concurrently examined the neural correlates of implicit knowledge about temporal order and chunk information, and provided converging evidence

of brain mechanisms that people acquire implicit knowledge about statistical regularities of temporal order based on the left parietal lobe, including the IPL and supramarginal gyrus, and scene chunks based on the left precuneus and bilateral PPA.

### **Brain activities to VSL during the familiarization scan**

In addition, we found neural evidence of VSL in the analyses of the familiarization scan. Specifically, brain activation in the left posterior cingulate cortex including the retrosplenial cortex was attenuated when participants viewed the structured blocks that contained temporal order information as compared to random blocks. In addition, the left precuneus and right caudate were more activated in the structured than in the random blocks depending on the difference in RTs between position 3 of the forward-foil and forward-triplet conditions. Thus, our findings are consistent with those of Turk-Browne et al. (2009) in suggesting that, based on the navigation studies (Hartley et al., 2003; Packard and McGaugh, 1996), the medial temporal lobe memory system may be abstract and it may generalize to new retrieval contexts, whereas the caudate memory system may be specific and require exact replication of an encoding context for learning to be expressed. It has been shown that the retrosplenial cortex as well as the PPA is related to scene perception, especially involving integrating viewpoints of scenes. Park and Chun (2009) assessed the activation in a panoramic condition where three natural images of different views from the same panoramic space were continuously presented; this was compared to an identical condition wherein the same set of three images was immediately repeated. Results indicated an impact of scene repetition in the retrosplenial cortex, and that this effect was eliminated in a non-continuous block where sets of the three images were not presented in immediate succession. Thus, our findings, together with previous ones, lead us to propose that the retrosplenial cortex is engaged in the extraction of regularities and connections of scene information (i.e., chunking) for representations that can be generalized to new contexts.

It should be noted that the pattern of our results differs from that reported by Turk-Browne et al. (2009). In particular, we observed repetition attenuation (i.e., lower activation of the posterior cingulate cortex in the structured versus random blocks), whereas they observed repetition enhancement. Recently, it has been reported that repetition enhancement and attenuation are observed differently as a function of familiarity of natural scenes: that is, the repetition of images leads to a more efficient representation of them compared to non-repeated images, resulting in repetition attenuation (Müller et al., 2013). Because we presented our participants with natural image regularities 12 times per each triplet, these stimuli may have become more familiar to participants, leading to a repetition attenuation in the structured versus random conditions.

**Conclusion**

In conclusion, the present experiment demonstrates the brain basis underlying implicit knowledge about triplet temporal order and scene chunks. In our daily life, we extract and learn statistical regularities based on categorical cues without awareness (Brady and Oliva, 2008; Otsuka et al., 2013), and the left posterior cingulate cortex and the right caudate are engaged in this type of learning. Using implicit knowledge about acquired regularities, people can efficiently detect target objects in natural scenes (Otsuka et al., 2016), and brain regions such as the left parietal lobe and the left precuneus are engaged in implicit knowledge about learned temporal order information and scene chunks.

### References

- Batterink LJ, Reber PJ, Neville HJ, Paller KA (2015) Implicit and explicit contributions to statistical learning. *J Mem Lang* 83:62–78
- Bertels J, Franco A, Destrebecqz A (2012) How implicit is visual statistical learning? *J Exp Psychol Learn Mem Cogn* 38:1425–1431
- Brady TF, Oliva A (2008) Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychol Sci* 19:678–685
- Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using an SPM toolbox. *NeuroImage* 16(suppl. 1):1140–1141
- Davenport JL, Potter MC (2004) Scene consistency in object and background perception. *Psychol Sci* 15:559–564
- Epstein RA, Higgins JS, Jablonski K, Feiler AM (2007) Visual scene processing in familiar and unfamiliar environments. *J Neurophysiol* 97:3670–3683
- Fiser J, Aslin RN (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci* 12:499–504
- Fiser J, Aslin RN (2002) Statistical learning of higher-order temporal structure from visual shape sequences. *J Exp Psychol Learn Mem Cogn* 28:458–467
- Hartley T, Maguire EA, Spiers HJ, Burgess N (2003) The well-worn route and the path less traveled: Distinct neural bases of route following and wayfinding in humans. *Neuron* 37:877–888
- Hatta T (1975) A study of handedness: Handedness and manual activity. *Tekisei-kenkyu* 9:1–13
- Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D (2013) The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain Lang* 127:46–54

Kim R, Seitz A, Feenstra H, Shams L (2009) Testing assumptions of statistical learning: Is it long-term and implicit? *Neurosci Lett* 461:145–149

Lancaster JL, Rainey LH, Summerlin JL, Freitas CS, Fox PT, Evans AC, Toga AW, Mazziotta JC (1997) Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Hum Brain Mapp* 5:238–242

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT (2000) Automated talairach atlas labels for functional brain mapping. *Hum Brain Mapp* 10:120–131

Li FF, VanRullen R, Koch C, Perona P (2002) Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci USA* 99:9596–9601

Lieberman MD, Cunningham WA (2009) Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Soc Cogn Affect Neurosci* 4:423–428

Maffei V., Giusti MA, Macaluso E, Lacquantini F, Viviani P (2014) Unfamiliar walking movements are detected early in the visual stream: An fMRI study. *Cereb Cortex* 25:2022-2034

Matsumoto D, Ekman P (1988) In Japanese and Caucasian facial expressions of emotion (JACFEE) and neutral faces (JACNeuF). San Francisco, CA: San Francisco State University.

Müller NG, Strumpf H, Scholz M, Baier B, Melloni L (2013) Repetition suppression versus enhancement: It's quantity that matters. *Cereb Cortex* 23:315–322

Otsuka S, Kawaguchi J (2007) Natural scene categorization with minimal attention: Evidence from negative priming. *Percept Psychophys* 69:1126–1139



- Otsuka S, Nishiyama M, Nakahara F, Kawaguchi J (2013) Visual statistical learning based on the perceptual and semantic information of objects. *J Exp Psychol Learn Mem Cogn* 39:196–207
- Otsuka S, Nishiyama M, Kawaguchi J (2014) Constraint of the semantic flexibility in visual statistical learning. *Vis Cogn* 22:865–880
- Otsuka S, Koch C, Saiki J (2016) Visual statistical learning produces implicit and explicit knowledge about temporal order information and scene chunks: Evidence from direct and indirect measures. *Vis Cogn* 24:155–172
- Packard MG, McGaugh JL (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol Learn Mem* 65:65–72
- Park S, Chun MM (2009) Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage* 47:1747–1756
- Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928
- Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB (2014) The necessity of the medial-temporal lobe for statistical learning. *J Cogn Neurosci* 26:1736–1747
- Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM (2013) Neural representations of events arise from temporal community structure. *Nat Neurosci* 16:486–492
- Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522
- Turk-Browne NB, Jungé JA, Scholl BJ (2005) The automaticity of visual statistical learning. *J Exp Psychol Gen* 134:552–564

- Turk-Browne NB, Simon MG, Sederberg PB (2012) Scene representations in parahippocampal cortex depend on temporal context. *J Neurosci* 32:7202–7207
- Turk-Browne NB, Scholl BJ (2009) Flexible visual statistical learning: Transfer across space and time. *J Exp Psychol Hum Percept Perform* 35:195–202
- Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK (2009) Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *J Cogn Neurosci* 21:1934–1945
- Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM (2010) Implicit perceptual anticipation triggered by statistical learning. *J Neurosci* 30:11177–11187
- Wan X, Nakatani H, Ueno K, Asamizuya T, Cheng K, Tanaka K (2011) Neural bases of intuitive best next-move generation in board game experts. *Science* 331: 341–346
- Xiao J, Hays J, Ehinger K, Oliva A, Torralba A (2010) SUN Database: Large-scale scene recognition from abbey to zoo. *IEEE Conference on Computer Vision and Pattern Recognition*

**Footnote**

<sup>1</sup>We also conducted anatomical ROI analyses on signal changes in both the familiarization and test scans according to previous studies (Schapiro et al., 2013; Turk-Browne et al., 2009): the left IFG (MNI coordinates,  $-28, 12, -6$ ), right caudate ( $8, 3, 9$ ), and right hippocampus ( $27, -27, -6$ ). However, we did not observe any significant effects of statistical learning.

**Acknowledgments**

This work was supported by JSPS KAKENHI [Grant Number 11J02236]. This study was conducted using the MRI scanner and related facilities of Kokoro Research Center, Kyoto University.

**Compliance with ethical standard****Conflict of interest**

The authors declare that they have no conflict of interest.

**Ethical standard**

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent**

Informed consent was obtained from all individual participants included in the study.

Table 1

*Results in the familiarization and test scans*

| <i>Structured blocks &lt; Random blocks</i>  |                   |          |          |          |               |               |
|--|-------------------|----------|----------|----------|---------------|---------------|
| <i>Region</i>                                | <i>Hemisphere</i> | <i>x</i> | <i>y</i> | <i>z</i> | <i>Extent</i> | <i>Peak t</i> |
| Posterior cingulate cortex                   | L                 | -12      | -56      | 16       | 264           | 5.18          |
| <i>Forward triplet &gt; Backward triplet</i> |                   |          |          |          |               |               |
| <i>Region</i>                                | <i>Hemisphere</i> | <i>x</i> | <i>y</i> | <i>z</i> | <i>Extent</i> | <i>Peak t</i> |
| Parietal lobe                                | L                 | -30      | -44      | 24       | 367           | 4.85          |
| <i>Both triplets &lt; Both foils</i>         |                   |          |          |          |               |               |
| <i>Region</i>                                | <i>Hemisphere</i> | <i>x</i> | <i>y</i> | <i>z</i> | <i>Extent</i> | <i>Peak t</i> |
| Precuneus                                    | L                 | -22      | -58      | 50       | 287           | 4.46          |

Regions exhibiting stronger responses to random versus structured blocks in the familiarization scan, to forward versus backward triplet blocks, and to both triplets versus both foils blocks ( $n = 24$ ; cluster level threshold,  $p < 0.05$  corrected);  $x/y/z$ , cluster center-of-gravity coordinates in MNI space; extent, number of suprathreshold voxels in the cluster; peak  $t$ , maximum statistic within the cluster; L: left.

Table 2

*Percent signal change as a function of each condition in the parahippocampal place area*

|         | Forward       | Backward      |
|---------|---------------|---------------|
| Triplet | 0.788 (0.070) | 0.808 (0.086) |
| Foil    | 0.829 (0.070) | 0.827 (0.083) |

Standard errors of the mean in parentheses.

### Figure legends

*Fig. 1.* Examples from the familiarization scan (2 structured and 2 random blocks). There were 24 (12 structured and 12 random blocks) 18-sec blocks, which included a rest period and a 800-ms digit display with a 200-ms ISI. Each block contained four structured, or four random triplets. In the structured blocks, three natural scenes consistently appeared in the same order within each triplet (e.g., ABC, DEF, GHI, JKL). In contrast, in the random blocks, the remaining 12 natural scenes were randomly presented in four random triplets with the following constraint: To maintain the same positional structure as structured triplets, each image always appeared in the same position within any of four random triplets. To wit, to control positional encoding, each scene appeared only in position 1 (e.g., A'-D'-G'-J'), position 2 (e.g., B'-E'-H'-K'), or position 3 (e.g., C'-F'-I'-L') of the four random triplets. Each of the twelve natural scenes in the block was presented individually for 800 ms followed by a 200-ms ISI. Then, a 5-sec rest period was inserted. In addition, one display containing three numbers (e.g., 364) was inserted in a random position after any one of the four triplets in each block, for 800 ms followed by a 200-ms ISI. During this scan, participants were required to detect the number display and press the button as quickly as possible. Solid-line squares reflect structured triplets, and dashed-line squares reflect random triplets.

*Fig. 2.* Examples from the test scan. On each trial, a target scene was presented for 2,000 ms, and participants were required to remember it. Following a fixation cross for 1,000 ms, five natural images were presented individually for 200 ms with an 800 ms-ISI. Of five images, the second (position 1 in each triplet or foil), third (position 2), and fourth images (position 3) consisted of triplets or foils. They were selected from the structured blocks during the familiarization scan. The first and fifth images were selected from the random blocks.

Participants were required to press the button with their index finger of the right hand as quickly as possible when they detected the target natural scene. The ITI was 4,000 ms.

*Fig. 3.* RTs in the target detection task of the test scan as a function of the triplet order (forward vs. backward), type of sequences (triplet vs. foil), and target position (position 1, position 2, vs. position 3). Error bars represent standard errors of the means.

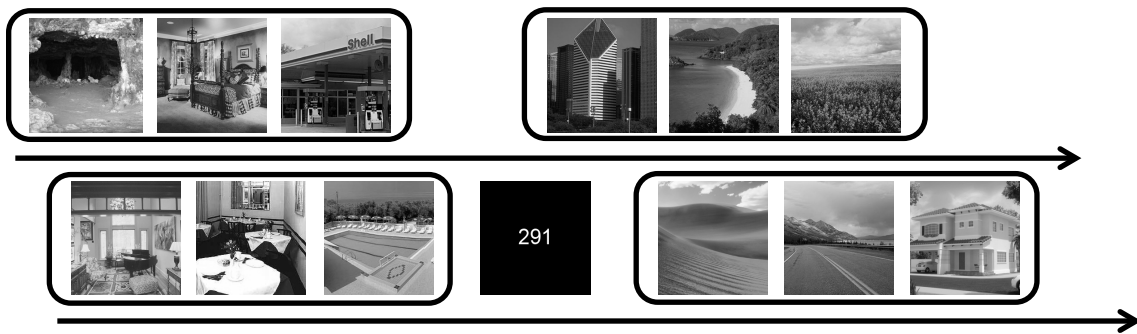
*Fig. 4.* Results of whole brain analyses in the familiarization scan. (a). The left posterior cingulate cortex exhibited stronger responses to random versus structured blocks (MNI coordinates,  $-12, -56, 16$ ;  $p = 0.041$  corrected; extent = 264; peak  $t = 5.18$ ). (b). Signal changes in the left posterior cingulate cortex of the structured and random conditions.

*Fig. 5.* Results of whole brain analyses in the test scan. (a). The left parietal lobe (MNI coordinates,  $-30, -44, 24$ ;  $p < 0.001$  corrected; extent = 367; peak  $t = 4.85$ ) exhibited stronger responses to forward versus backward triplets. (b). Signal changes in the left parietal lobe of the forward- and backward-triplet conditions.

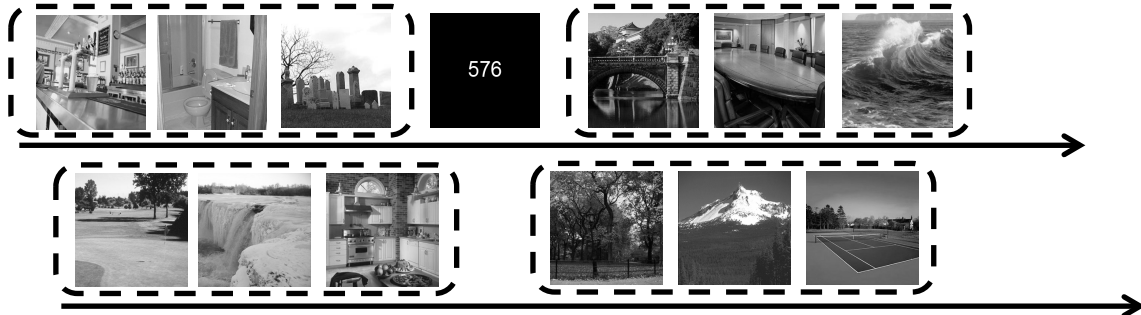
*Fig. 6.* Results of whole brain analyses in the test scan. (a). The left precuneus exhibited stronger responses to triplets (i.e., forward triplet + backward triplet) versus foils (i.e., forward foil + backward foil) in the test scan (MNI coordinates,  $-22, -58, 50$ ;  $p = 0.011$  corrected; extent = 287; peak  $t = 4.46$ ). (b). Signal changes in the left precuneus of the triplet (i.e., forward triplet + backward triplet) and foil conditions (i.e., forward foil + backward foil) during the test scan.



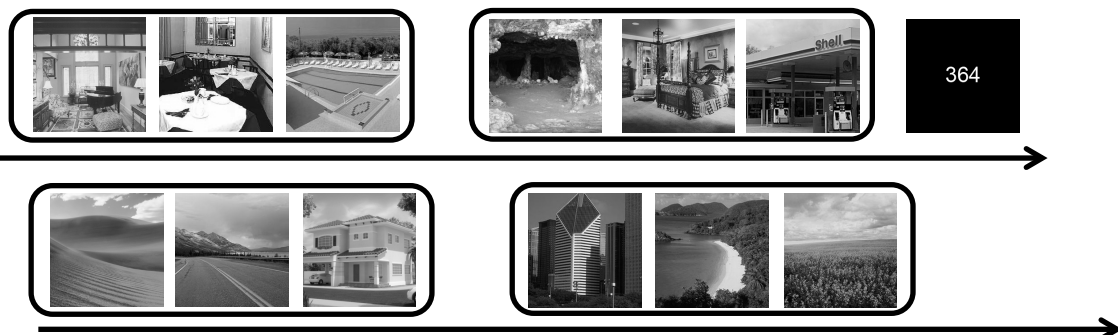
Structured block



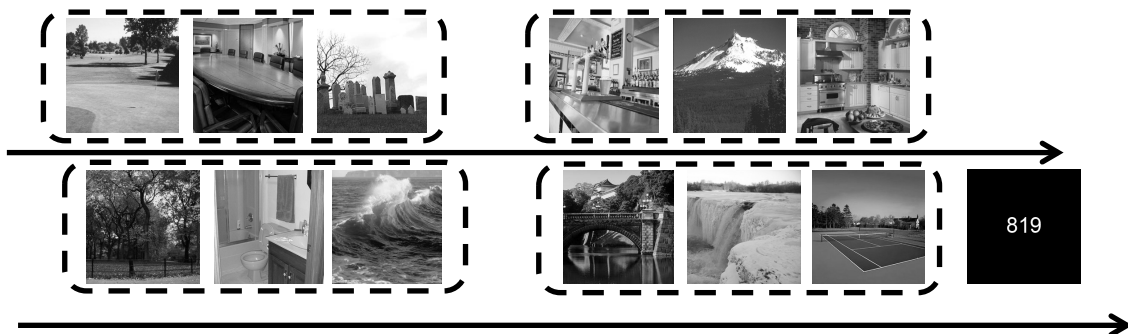
Random block

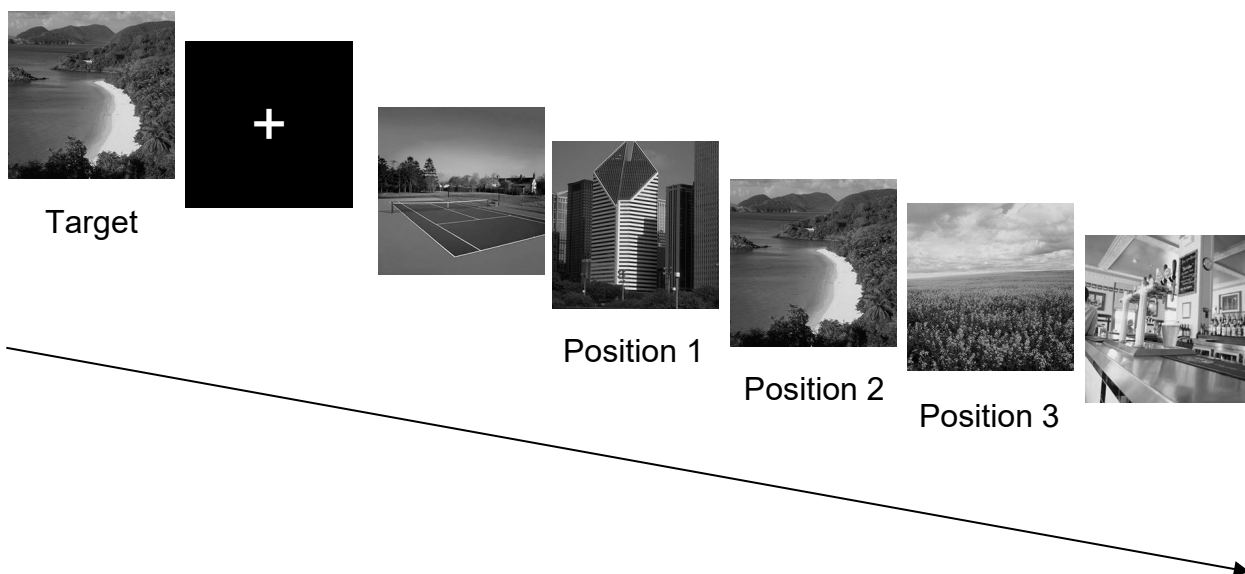


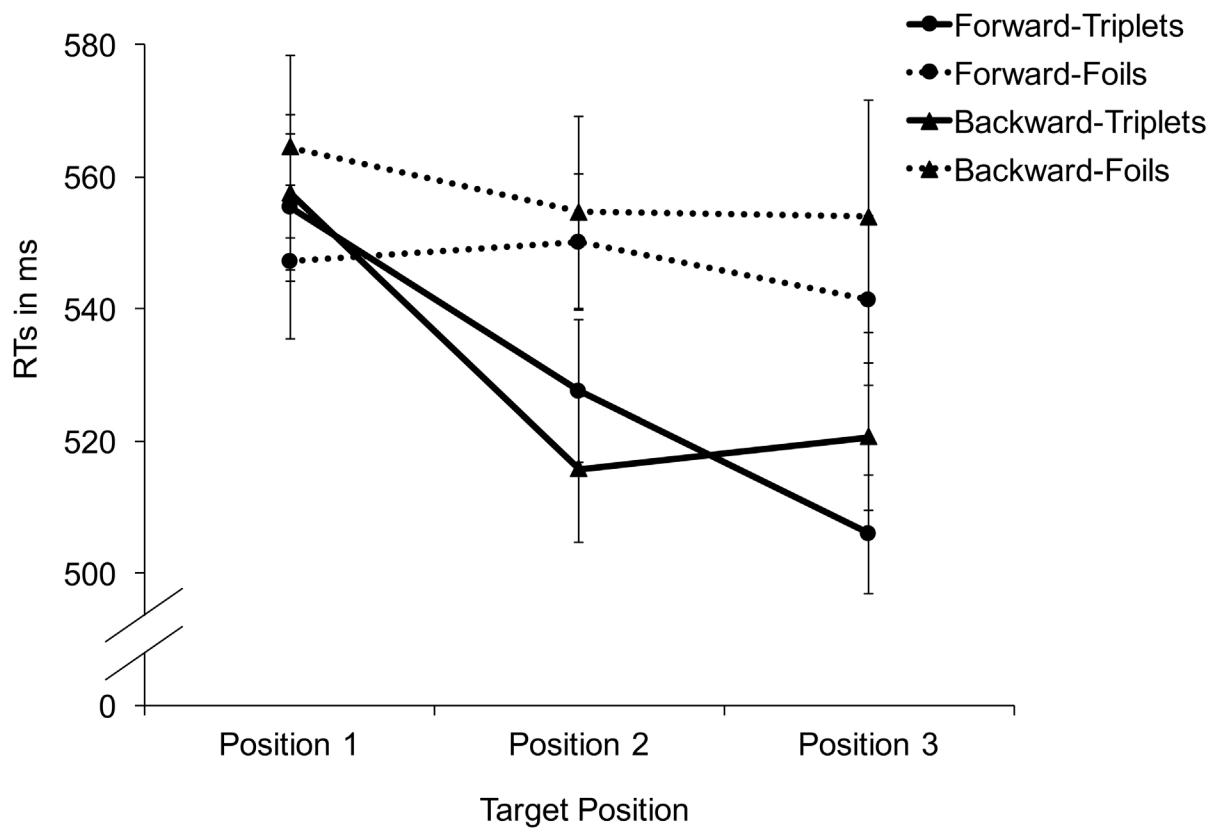
Structured block



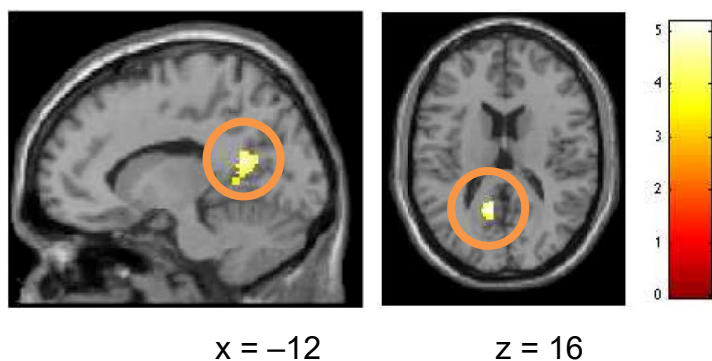
Random block



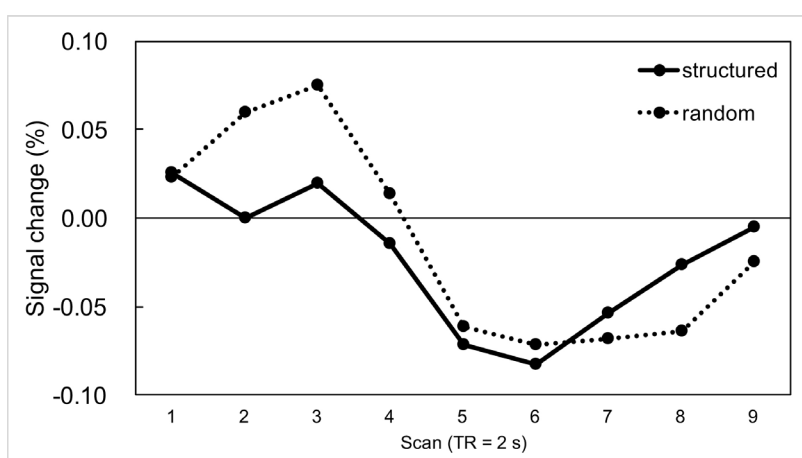




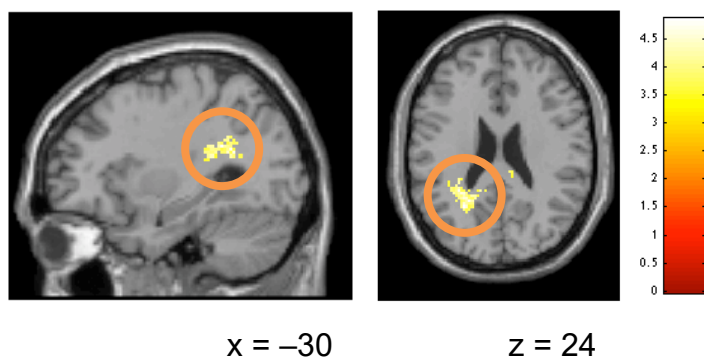
a.



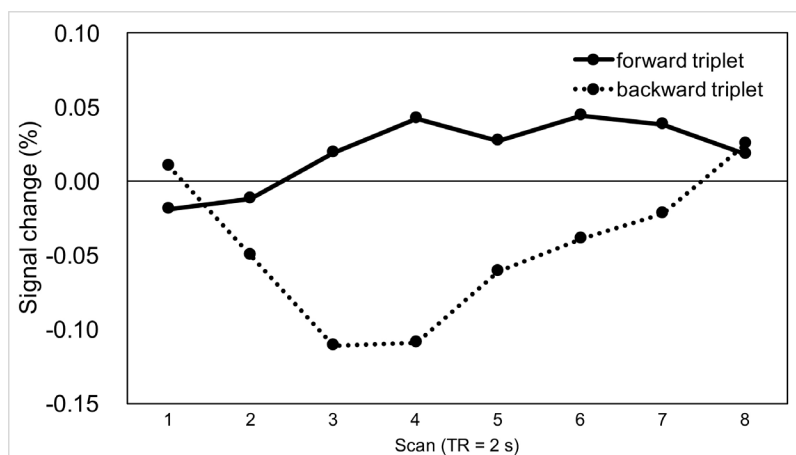
b.



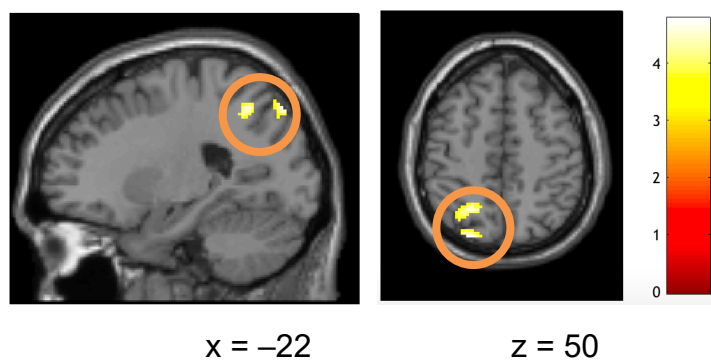
a.



b.



a.



b.

