An EEG-based neural decoding approach for investigating statistical learning between modalities

Daniel Duato Catalán

Master of Artificial Intelligence
Facultat d'Informàtica de Barcelona
Universitat Politècnica de Catalunya

Tutor:	Dr Alfredo Vellido Alcacena
Advisors:	Dr Alexis Pérez Bellido
	Dr Ruth de Diego Balaguer

October, 2018

Abstract

Statistical learning (SL) is the human ability to extract statistical regularities from the environment. Most of the previous studies on SL have focused on the regularity extraction mechanisms taking place within sensory modalities but there is little evidence that statistical regularities can be learnt across different modalities. To test whether SL is a domain-general or modality-specific mechanism (Frost et al., 2015), we conducted two experiments in which participants where exposed with a stream of visual and auditory non-meaningful stimuli whereas they performed an oddball detection task. Stimuli were grouped into unimodal (*VV* or *AA*) or crossmodal (*VA* or *AV*) pairs and the only cue to identify a pair was a higher transitional probability between the paired elements.

In Experiment 1 (15 participants), we measured learning using a 2AFC task and gathered the participants confidence on their answers. We found that only the unimodal transitional probabilities could be explicitly reported above chance level. In Experiment 2 (2 participants) we used electroencephalography (EEG) to obtain neurophysiological measures of learning. We investigated event-related potentials (ERPs) and used neural decoding with the temporal generalization method to investigate stimulus preparation.

In general, our results favor the hypothesis that SL is a domain-specific system tuned to work within but not between sensory modalities.

Acknowledgements

First I would like to thank my tutor Alfredo Vellido for sparking my interest in computational neuroscience and making me start this adventure. Special thanks to my advisors Alexis Pérez and Ruth de Diego, for being so helpful through all this process and providing me with so many opportunities.

Finally thanks to the whole Brainvitge Cognition & Brain Plasticity team, where I've been working on this dissertation for the past year, for making me feel at home since the beginning. I really can not imagine a laboratory with better colleagues.

Contents

Chapter 1

Introduction

The information present in our environment is generally organised conforming regular patterns. In this sense, most of the objects that we perceive can help us forecast the presence or appearance of other objects in time and space. As humans, our brains have the ability to automatically learn these patterns without being aware of their learning. For example, imagine someone going to work every day following the same path, without paying explicit attention to the buildings surrounding her. However, one day, one of these buildings is demolished and this person notices the change, to her surprise, since she was not aware of having learnt the pattern of buildings. Through repeated exposure to the patterns, she had learnt them automatically. This automatic learning of patterns just through exposure to them has been called statistical learning, since it consists of sensing, extracting and learning the statistical regularities from our environment. This mechanism has been characterized, for instance, as an important component in language acquisition by infants, particularly in word segmentation. It is in this field that the term was coined [1] and where much of research on statistical learning has been focused on [2][3].

The study of statistical learning consists of exposing experimental subjects to a collection of stimuli that have controlled statistical regularities and later testing subjects' knowledge about such regularities. Of course, this definition encompasses many different types of learning and of statistical regularities, but statistical learning research has mainly focused on one of those, namely the transitional probabilities (TP). These are the probabilities that one object will be followed by another one every time this object appears. They are calculated by dividing the apparition frequency of the sequence of both objects by the apparition frequency of the first object. Humans seem to be more sensitive to these statistics than to other simpler ones, like frequency count.

For instance, let us examine the word segmentation example. When we speak, we do not always introduce noticeable pauses in between words that would enable a naive listener to detect where words end and where they begin. A statistical learner could anyway identify word boundaries by computing statistical regularities between syllables. If the TPs between two syllables is high, they most probably are part of a word, since they appear often together. If the TPs are low, most probably they belong to two separate words and a word boundary may be identified at this position.

Examples of this type of learning have been found regarding many different types of objects besides language and in different sensory modalities. And not only in vision (as in our previous example about the buildings) or audition (as in language acquisition), but even in the tactile modality. Given that these processes work in very similar ways on the different modalities, a natural question is whether a supramodal mechanism governs the statistical learning in all sensory modalities, or whether statistical learning is a modular process entailing separate cognitive mechanisms, each one encapsulated within their modality.

Experiments exploring the relationships between modalities during statistical learning processes have been carried out using multiple approaches. For example, learning patterns of multimodal objects, transferring knowledge from one modality to another, having stimuli from one modality affect the learning of another modality pattern [4] and simultaneously learning patterns from different modalities[5]. All these have been proven to work, adding to the body of evidence of integration between modalities in statistical learning while also revealing some of the constraints between them.

However, as far as we know, it has not been researched whether humans also learn task-unrelated cross-modal sequential patterns automatically, i.e., patterns that consist of elements of different modalities that can predict each other. For this to be possible, we should be able to encode TPs between elements of different modalities.

We know that it is possible to learn dependencies between objects of different modalities, as for example in classical conditioning. In the most famous example of classical conditioning, a dog is able to learn, through repeated exposure, a dependency between the sound of a bell and the apparition of food, therefore associating an auditory stimulus with a visual or olfactory one [6]. These types of experiments have been widely replicated also with humans, for example by making children associate furry objects with harsh sounds [7].

These experiments deal for the most part with high-level objects that are already meaningful to the subjects, which are explicitly confined in time and space and have associations related to the subjects' survival that act as strong external incentives for learning (food is required not to starve and harsh noises can be indicators of danger). Given an already meaningful object, our brains can direct our attention towards its temporal and spatial neighbourhood in order to find patterns that can teach us more about the nature of the object and help us predict it better in the future, even if these patterns are comprised by objects of different modalities. Some experiments have shown the possibility of learning dependencies between a priori non-meaningful stimuli of different modalities when the dependencies are directly related to the task of the participant, thus rendering those stimuli and dependencies meaningful for the specific context [8].

However, this learning process seems to be different from the automatic detection of patterns between non-meaningful objects associated with statistical learning, such as in the learning of artificial languages, where there are no clear external incentives or temporal boundaries that may help to direct our attention.

1.1 Objective

In this work we will specifically address the question whether humans can automatically learn the TP between non-meaningful task-unrelated cross-modal objects in order to dissociate between two different hypotheses:

1. SL is a domain-general cognitive mechanism that can integrate multiple modalities and extract statistical regularities between them.
2. SL is a domain-specific cognitive mechanism that can only extract statistical regularities between objects of the same modality. Associations between objects of different modalities are formed only through higher-level cognitive mechanisms once they become meaningful objects.

1.2 Experimental paradigm

In our experiment, we used the common SL paradigm, in which participants are exposed to a stream of non-meaningful stimuli that has some hidden structures, unbeknownst to the participant, and then tested about their knowledge of such structures.

We wanted to start with the simplest possible structure, which would be adjacent temporal dependencies between stimuli, forming a set of stimuli pairs. We wanted to test cross-modal dependencies in both directions: auditory to visual and visual to auditory. In addition to this, we wanted to include unimodal dependencies, so that we can both check the validity of our paradigm and materials, given that we would expect unimodal learning to work on the basis of the literature; and use this type of learning as a baseline to compare cross-modal learning against. This results in four different types of dependency in our experiments: auditory to auditory (*AA*), visual to visual (*VV*), auditory to visual (*AV*) and visual to auditory (*VA*).

We considered separating the conditions into different blocks. However, that would add a repetitive predictable pattern of modality switches to the cross-modal conditions (*AVAVAV* or *VAVAVA*) that could either interfere with the statistical learning mechanism by making both modalities more easily differentiable or enhance the detection of patterns by making their periodicity more apparent. Even after solving this rhythmic issue, for example by adding noise stimuli that are not attached to any dependency, comparison between unimodal and cross-modal conditions would still be unreliable because of differences in the modality switching effect. Besides, a participant can only be tested for one block, since we do not want them to know about the test while they are being exposed to the stream. So we decided to group all four conditions into one single stream.

Another requirement of our experiment was that we wanted participants to drive their attention towards the stream of visual and auditory attention, but did not want to explicitly drive their attention towards the hidden statistical regularities. To do this we included an oddball detection task which was orthogonal to the hidden structure of pairs. So, participants had to detect oddball stimuli (both visual and auditory) that randomly appeared during the stream, but the hidden statistical regularities were completely uninformative about the position. Therefore, learning of the statistical regularities could be considered to be task-unrelated.

The inclusion of this task also served an additional purpose, namely to use the reaction time to oddballs as an online measure of learning. This is based around the concept of stimulus preparation, which is the preparation of the neuronal networks required for encoding a stimulus. This preparation happens when the stimulus is being predicted. Therefore, if a stimulus is being predicted due to a dependency having been learnt, we would expect the time required to encode the stimulus to be reduced; which would also reduce the total time required to discriminate whether the stimulus is an oddball or not. For this measure to work, oddball and standard stimuli needed to share similar stimulus preparation activations and encoding processes, so we used slightly distorted versions of the standard stimuli as oddball stimuli.

However, we did not finally include this measure in our analysis due to a bug in the experiment program. The bug caused some of the keypresses to be registered as happening at a slightly different time than their true occurrence. This shift in the registered times was small enough (20 ms at maximum) that it did not affect any of our others computations relating oddball stimuli, but it caused any results of this analysis to become unreliable.

Two versions of this experiment were conducted, each one with a different approach to measure learning effects.

In experiment 1, we assessed whether participants had learnt or not through a behavioural test after they had been exposed to the stream of stimuli. The test was a two-alternative forced choice task, in which the effects of learning were measured by the capacity of participants to, given two pairs of stimuli, identify which one shared statistical regularities with the stream of stimuli they had been previously exposed to. This is a common test used in SL research to assess learning [9][2]. We were also interested in knowing how explicit or implicit (i.e., whether the person is aware or not of possessing such knowledge) cross-modal learning is compared to unimodal learning. This question was of interest since, if the domain-specific SL hypothesis were true, we might expect that a higher threshold of explicitness was required for cross-modal dependencies to be learnt than for unimodal ones. So, if the correctly identified cross-modal pairs were, by average, more explicitly recalled than the correctly identified unimodal ones, we could attribute this cross-modal learning to cognitive processes of higher-level than unimodal SL, hence favouring the second hypothesis. To test this, we asked participants after each test trial to report their degree of confidence about their previous answer using the same procedure as [10].

In Experiment 2 we investigated learning through internal neurophysiological markers of learning using electroencephalography (EEG). We analyzed brain activity during the exposure to the stream of stimuli using two approaches. In the first place, we compared event-related potentials (ERPs) in response to stimuli with different levels of expectancy. If learning was taking place we expected to see, in comparison to brain activity response stimuli unaffected by predictions, an expectation suppression effect in the response to predictable stimuli and a surprise enhancement effect in the response to stimuli that violate a prediction. These effects have been used as markers for learning in previous SL research on both humans [3][11] and animals [12][13].

Secondly, we looked for neurophysiological markers of stimulus preparation preceding a predictable stimulus. The brain activity of stimulus preparation appears to be similar to the one of the stimulus that is being predicted [14]. So, in order to identify this effect, we measured up to which point the brain activity preceding a predictable stimulus shared information about the stimulus by training neural decoders through Linear Discriminant Analysis (LDA) on the brain activity during a stimulus and testing it on the period preceding the stimulus. Training decoders on one timepoint and testing it on other timepoints is called temporal generalization and has been used to identify similarities between mental representation and get a better picture of the temporal dynamics of cognitive processes [15][16].

With these two neurophysiological analyses we hoped to notice effects of implicit learning that may slip past the explicit behavioural test, since the recollection prompted by the test could interfere with implicitly learnt knowledge.

Chapter 2

Methods

In this chapter, we provide a self-contained account of some of the methods employed in the experiments of the thesis. This includes details on data acquisition methods, such as Electroencephalography (EEG) and their analysis, such as Event Related Potentials (ERP). It also include descriptions of data analysis methods such as Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA).

2.1 Electroencephalography (EEG)

EEG is a non-invasive neuroimaging technique that can measure the electrical activity from the brain placing electrodes along the scalp. The electrical activity measured is the one produced by the excitatory or inhibitory postsynaptic potentials of synchronized clusters of similarly oriented neural cells, mostly pyramidal cells in the cerebral cortex. The voltages recorded are potential differences between a reference electrode and the scalp electrode we are interested in measuring, referred to as channel electrode.

This neuroimaging technique has very low spatial resolution, since the electric signals get distorted when passing through the dura mater, skull and scalp. The folds of the cortex and signals bouncing against the skull can make the electric signals interfere with one another and highly displace the location where a signal is measured with respect to where it originated. There are methods for source localization that tackle this problem, but in general the spatial resolution of EEG is very low and one should avoid thinking of an EEG topography as a direct representation of the spatial location of brain activity.

On the upside, EEG recording possesses a very high temporal resolution, up to the milliseconds. This high temporal accuracy and its non-invasiveness have made it very popular in neuroscience research.

2.1.1 Event-related potentials (ERPs)

When analyzing EEG recordings, a researcher may be interested in the brain activity surrounding a certain event, such as the onset of a stimulus or a keypress. However there is very high variability between trials, due to biological and external artifacts or to the different thoughts that the subject has at any moment. To solve this problem, a high number of samples of activity around the event of interest are needed, so that they can be averaged in order to better approximate the true signal of interest. The process of cutting a continuous EEG recording into samples (or epochs) around a certain event is called timelocking. The averaged signal of those epochs is what we know as ERP. It is convention to plot ERPs with the potential axis flipped.

An ERP curve is made up of a set of positive and negative peaks. These curves and peaks are the composition of electrical activities relating to different brain functions. Research of ERPs centers around relating these curve characteristics to brain function through the measure of differences in them (e.g., of the peaks' amplitude or latency) given variations in experimental conditions. Positive or negative peaks that are associated to a certain temporal period and cognitive processes are called components. They are often referred to by a letter, which indicates it's polarity (either N or P), and by a number that indicates its usual temporal position. For example, the component N400 component is a component with negative amplitude that appears around 400 ms after the event to which the ERP is timelocked.

In this study we looked for components that have been associated with the effects expectation suppression and surprise enhancement, which can be used to identify learning. Expectation suppression is the reduction in amplitude of components produced by a the prediction of a stimulus. Surprise enhancement is the increase in amplitude of components when a prediction is violated.

2.2 Independent Component Analysis

ICA is signal separation technique used to identify the source signals of a multivariate signal that assumes non-Gaussianity and independence of the source signals. In mathematical terms, given a multivariate signal $\boldsymbol{X}$, composed of data samples $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ with n features, the task of ICA is to find a transformation matrix $\boldsymbol{W}$ comprised of weights $w_{i,j}$ that can project the data into a new space $\boldsymbol{Y}$, with a new set of components $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)$, such that $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$, where $y_i = (w_{i,1}x_1 + w_{i,2}x_2 + \cdots + w_{i,n}x_n)$, and optimizing an independence function $g(\boldsymbol{Y})$.

It can be used in neuroscience research to identify ERP components or during signal preprocessing in order to remove artifacts such as blinks, eye movements or heartbeats [17][18]. It is particularly useful for the later, since the signal produced by these artifacts is highly independent from other sources of activity, so they are easily separated into distinct components when applying ICA. Once separated, they can be identified for example by looking at their ERPs or topography (which corresponds to their weights $w_{i,*}$). Once the artifact components have been identified, they are discarded (i.e., their values y_i are set to zero) and the remaining components are projected back into the original signal's space by multiplying them with the inverse of the weights matrix $\boldsymbol{W}^{-1}$.

There are multiple approaches to compute this, and the one we used in this study was infomax-based ICA[19]. This method is based on the maximization of the entropy of $\boldsymbol{Y}$, thus maximizing the independence between its features by maximizing the mutual information between $\boldsymbol{Y}$ and $\boldsymbol{X}$. This mutual information is defined as $I(\boldsymbol{Y}, \boldsymbol{X}) = H(\boldsymbol{y}) - H(\boldsymbol{Y}|\boldsymbol{X})$, where $H(\boldsymbol{Y})$ is the entropy of $\boldsymbol{Y}$ and $H(\boldsymbol{Y}|\boldsymbol{X})$ is the entropy of $\boldsymbol{Y}$ conditioned on $\boldsymbol{X}$, which intuitively means how much information is contained in $\boldsymbol{Y}$ that was not included in $\boldsymbol{X}$. In a model that does not contemplate noise (as the one used in common ICA, where noise is usually treated as another independent component), this conditional entropy can be neglected, since the relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is simply the linear transformation matrix $\boldsymbol{W}$. So, $I(\boldsymbol{Y}, \boldsymbol{X})$ and $H(\boldsymbol{Y})$ become directly proportional. To maximize $I(\boldsymbol{Y}, \boldsymbol{X})$, the relationship between the probability density functions of $\boldsymbol{Y}$ and $\boldsymbol{X}$ is traversed using stochastic gradient ascent. This gradient ascent corresponds to updating $\boldsymbol{W}$ by maximizing its Jacobian.

We used the *runica* implementation of this method, included in the EEGLab Toolbox for Matlab [20], for artifact removal during EEG pre-processing.

2.3 LDA

LDA is a supervised dimensionality reduction technique that can be used to obtain linear classifiers of multivariate data. Given a set of multivariate data $\boldsymbol{X}$, where each sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ has n features and belongs to a known class, LDA works by finding a set of weights $\boldsymbol{W}$, so that its eigenvectors project the data into a space that maximizes between-class variance while minimizing the within-class variance of the projected data. More precisely, it is Fisher's criterion what is being maximized:

$$\arg\max_{W} \frac{\boldsymbol{W}^T S_B \boldsymbol{W}}{\boldsymbol{W}^T S_W \boldsymbol{W}} \tag{2.1}$$

where S_B is the between-class covariance matrix and S_W the within-class covariance matrix. The weight matrix that maximizes this criterion can then be calculated as follows

$$\boldsymbol{W} = S_W^{-1} S_B \tag{2.2}$$

Once $\boldsymbol{W}$ has been calculated, we obtain its eigenvalues $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ and corresponding eigenvectors $\boldsymbol{V} = (v_1, v_2, \ldots, v_n)$. Each eigenvector can be used to project the data into a new dimension and their corresponding eigenvalues represent the discrimination power of the projection; the higher the eigenvalue the higher the between-class variance and the lower the within-class variance of the projection. So, if the objective is to reduce the dimensionality of the data down to k features, the eigenvectors can be sorted in descending order according to their eigenvalues and the first k of them selected, to form a transformation matrix that can project the data from a n-dimensional space to a k-dimensional one. The results from the LDA can also be directly used as a linear discriminator, by selecting only the eigenvector with maximal eigenvalue, since it's projection maximizes the separability of the data.

A problem of LDA occurrs when dealing with a very high-dimensional data sample, which can lead to a singular S_W matrix that makes its inversion and the computation of $\boldsymbol{W}$ impossible. There are different approaches to deal with this problem, one of them being shrinkage (or regularization) of the covariance matrix [21]. This method softens extreme values that may appear in the within-class covariance matrix according to a parameter γ. The optimal value of γ in order to increase discriminability can be analytically calculated as explained in [22].

We did not face small-sample size problems, since our sample size was bigger than the number of features (29); however, we still performed shrinkage regularization given the improvements in performance reported by applying it to neural decoding tasks [23]. The optimal γ value we obtained was very close to zero, producing almost no regularization and showing that it was not necessary in our case.

2.3.1 Temporal generalization

In our neural decoding analysis we used an approach called temporal generalization. This method deals with a collection of trials belonging to some known classes, where each trial has both a feature dimension consisting of n features and a temporal dimension consisting of m timepoints and therefore each trial actually consists of $m \times n$ values. For example, a trial could be the activity of all electrode channels in an ERP. This method serves as a way of measuring the information correlation between two timepoints. It consists of training several classifiers (also referred to as decoders) to discriminate class, each one with data of a different timepoint, and then testing them in all timepoints. This results in a matrix of accuracies $\boldsymbol{A} \in \mathbb{R}^{m \times m}$ with the dimensions *training time* and *decoding time*. The diagonal of this matrix corresponds to normal classification of a single timepoint and so it indicates the information about the class contained in that timepoint. Points outside of the diagonal show how well a decoder trained on a data from a timepoint can generalize to data from a different timepoint, thus showing that the class information in similarly encoded in both timepoints. For example, high accuracy in the position $a_{t,d} \in \boldsymbol{A}$ would indicate that class-related information is similarly encoded during timepoints t and d. A low accuracy (below chance level) would indicate an inverse correlation between the class encoding at different timepoints.

This approach can be very useful to explore the spatiotemporal dynamics of cognitive processes during ERPs, since similarities in encoding at different timepoints can indicate that the same (or similar) cognitive processes are active at those timepoints.

Chapter 3

Experiments

In this chapter we will present both conducted experiments, whose main experimental paradigm has been described above. For each experiment we will describe the materials used, the procedure followed and the analysis carried (including the preprocessing steps in the case of the EEG experiment). A brief discussion follows the exposition of the results for each experiment.

3.1 Experiment 1: Behavioural experiment

3.1.1 Participants

Twenty-eight naive adults participated in the experiment in return for remuneration. Before the study, participants provided written informed consent. They reported no hearing problems and normal or corrected-to-normal visual acuity. We excluded from the analysis participants that failed to detect more than 75% of oddballs in any modality (7 participants) and participants that had more false detections than correct ones (3 participants). We also excluded 3 participants due to experimental errors. Finally, fifteen participants (mean age of 22.9 years, 9 female and 6 male) were included in the analysis.

3.1.2 Materials

Visual stimuli consisted of artificially generated images of symmetric round fractals of 800 px of diameter. In total, 100 images were created in Apophysis 7x, a fractal flame renderization software that uses iterated function systems [24]. The procedure consisted of building a function system out of random linear and nonlinear transformations, forcing the resulting fractal to adopt a cyclic symmetry of order from 3 to 12, rendering it with a random color palette and filtering the image through a round Tukey window that faded the borders of the fractal into transparency. We manually selected 12 of them trying to maximize their perceptual dissimilarity. The luminance saliency of the selected stimuli was normalized using histogram equalization applied to the value channel of their HSV (hue, saturation, value) representation. Visual stimuli were presented in the middle of a CRT monitor, above a black background. The viewing distance from the participants to the monitor was approximately 50 cm. The stimuli were downscaled as needed in order to occupy 20^{o} of the participants visual field.

Auditory stimuli were created using random parametrization of multiple synthesizers from the digital audio workstation FL Studio. Sounds were generated this way until we got 12 differentiable enough stimuli. They were then edited using the software Audacity. Their amplitude was normalized and their length reduced to 500 ms by cutting and time-stretching. To remove any clicking in the sounds we applied to them an envelope of 100 ms of attack and 100 ms of release. Auditory stimuli were presented through speakers located at the sides of the monitor, to avoid the addition of a spatial attention shift effect when modalities switched.

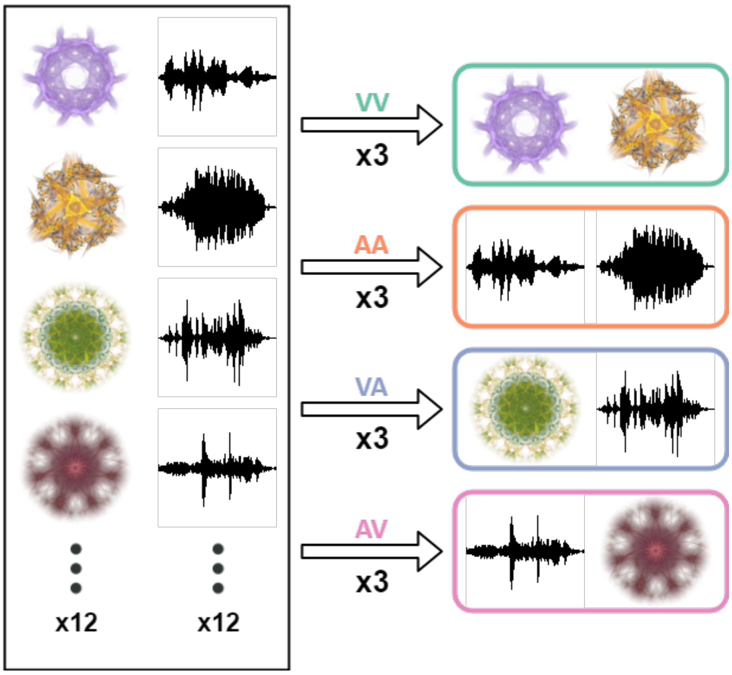

Figure 3.1: Illustration of the creation of standard pairs. Out of 12 stimuli of each modality, 3 pairs were created for each pair modality.

Those 24 stimuli were randomly grouped for each participant into 12 ordered standard pairs: three pairs for each one of the four pair conditions. During exposure, pairs were presented sequentially to the participant, ordered in such a way that any pair was followed an equal amount of times by any other pair. The sequence was created by finding an Euler tour in a directed graph whose vertices represent the transitions between different pairs. This way transitional probabilities between the stimuli within a standard pair always stayed at 100% and the ones between stimuli of different pairs at around 9%. Each pair was presented 110 times during exposure. Three different pair sequences were created and they were counterbalanced across participants.

Oddball versions of all stimuli were created for the attentional task. The oddball versions were designed so that they could be easily discriminated from the standard stimuli while still sharing enough characteristics with their corresponding standard stimulus so that they could be identified as alterations of each standard stimuli and would share similar encoding processes with them. In the visual modality, a layer of upscaled RGB white Gaussian noise with the same shape as the stimuli was added to each stimuli. The noise was upscaled in order to get an additional pixelated effect. In the

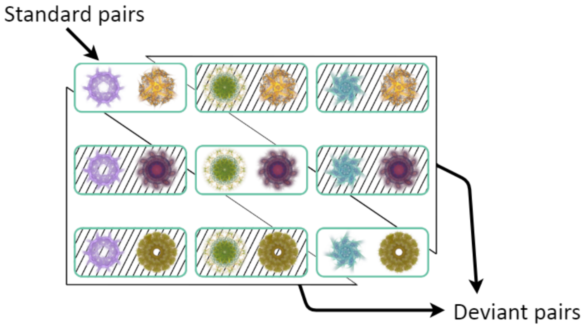

Figure 3.2: Illustration of the creation of deviant pairs for the test trials of Experiment 1.

auditory modality, white Gaussian noise with the same amplitude as the stimuli was added to the stimuli. Each stimulus appeared in its oddball version 10 times, out of a total of 110. These oddball appearances were randomly scattered across the exposure phase.

After exposure to the sequence of auditory and visual objects, participants were tested for their learning of the contingencies. For this test phase, we created six deviant pairs for each pair condition used in the exposure, by replacing the trail stimulus of the three standard pairs by the trail stimulus of each one of the two remaining standard pairs of the same pair condition. Figure 3.2 illustrates this process.

3.1.3 Procedure

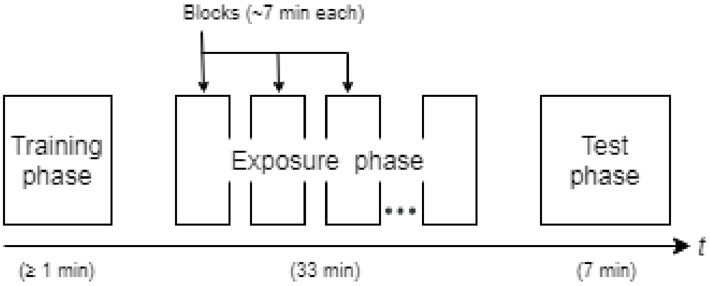

Figure 3.3: Diagram of the phases of Experiment 1.

The experiment consisted of three phases: a training phase, in which participants got familiar with the stimuli and the task; an exposure phase, in which participants were exposed to statistical regularities; and a test phase, in which participants were tested regarding their knowledge of such regularities. Each phase was separated by a short break.

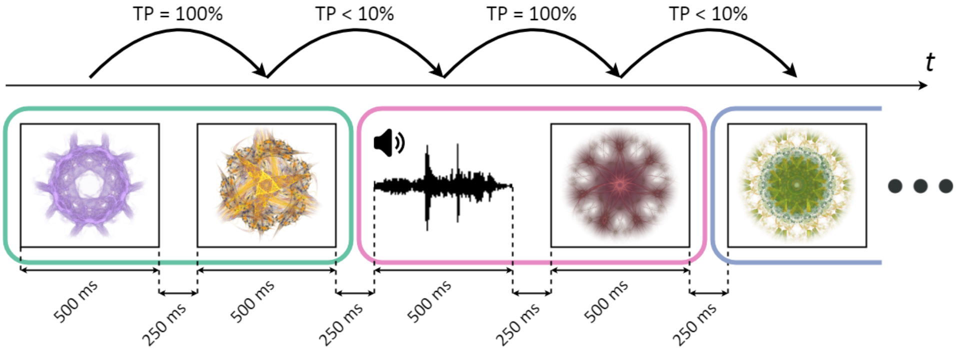

Figure 3.4: An example of a segment of the exposure phase, showing the difference in TPs depending on the pairs structure.

During the training and exposure phases, and in order to keep the participants' attention in both modalities, participants were asked to perform a cover task orthogonal to the objectives of the study in which they had to detect oddball stimuli and press the spacebar key whenever they saw or heard one. They were instructed to direct their gaze at a white fixation cross at the center of the monitor. Instructions were given to the participants both in written form printed on the screen and orally by the experimenter.

One example of a stimulus per modality and their corresponding oddballs were shown to the participants during the instructions.

During the training and exposure phases, each visual or auditory stimulus was presented for a duration of 500 ms separated by a 250 ms black interval.

During the training phase, stimuli were presented to the participant in a random order, without any regard to the pair structures. During this phase, oddballs appeared at a higher rate than during the exposure phase (35% of the stimuli), because the objective of this phase was to teach participants to discriminate between normal and oddball stimuli. The training phase lasted for one minute, but was repeated as needed until the experimenter was confident that the participant had understood the task and learnt to differentiate the oddball stimuli.

In the exposure phase participants were exposed to the sequence of standard pairs while being unaware of its underlying structure and while performing the cover task. This phase was divided into 5 blocks, each of them lasting for approximately 6 and a half minutes. Between blocks, participants were given the chance to rest and then resume the experiment by pressing a key whenever they wanted. After each block, participants received feedback on the monitor about the percentage of oddball stimuli detected, in order to keep them engaged in the task during the experiment. For this computation, we considered an oddball stimulus as detected whenever a keypress occurred no later than 2 seconds after the stimulus' onset.

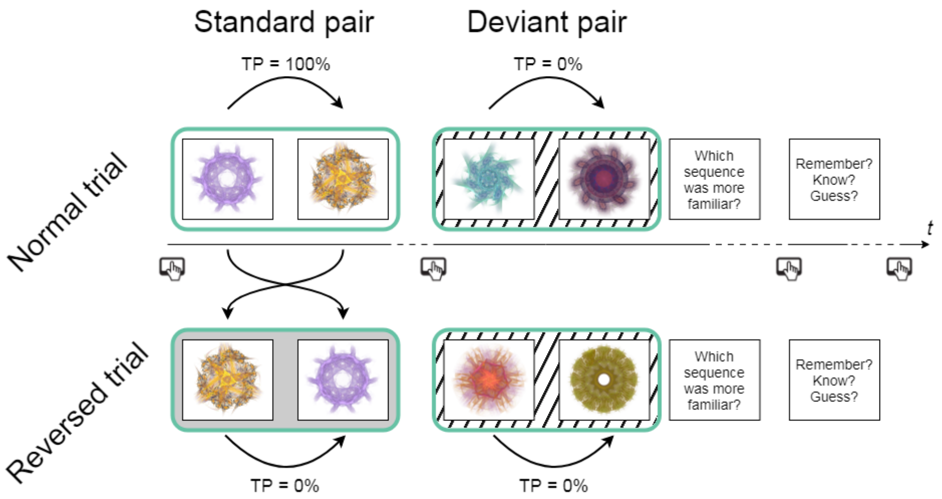

Figure 3.5: Example of both a *Normal* trial and a *Reversed* trial, showing the transitional probabilities (TPs) between the stimuli of the pairs.

After the exposure phase, we revealed for the first time to the participants that there is an additional task they had to undertake. During the test phase participants had to perform a two-alternatives forced choice task (2AFC). Each trial consisted of two stimuli pairs: a standard one, meaning that it shared statistical regularities with the stimuli sequence of the exposure phase; and a deviant one, meaning that it violated such statistical regularities.

Besides the *Normal* trials, we included an additional type of trial in the test, *Reversed* trials. In *Normal* trials, the familiar pair was one of the standard pairs and the deviant pair maintained the correct order of presentation of objects in relation to their position in the standard pairs. However in the Reversed trials, the stimuli of both standard and deviant pairs was presented in the reverse order (i.e., the trailing stimulus appearing first and the leading stimulus appearing last). We consider the reversed standard pair as the familiar pair during *Reversed* trials, since the association between stimuli is the same as in the standard pairs.

We included *Reversed* trials in the test for two reasons. First, they would serve as ‘noise’ trials that would make difficult for participants to use explicit deductive strategies that could help them to correctly recognize all standard pairs. This strategy can follow from the assumption that in each trial there was a correct and a false answer and just a single trial in which they were certain about their answer (so it could also make them incorrectly recognize all deviant pairs, in case the participant was wrong in their certainty). Secondly, it also served to test whether, in the case of learning, participants were only learning TPs (which should be order-sensitive) or some other order-insensitive association. If participants learned the association irrespective of the order, we expected that they should perform better than chance (i.e. >50% responses) at recognizing the familiar pair in both *Reversed* and *Normal* trials. If they learned TPs with the specific order, participants should only perform better than chance in the Normal trials and if they did not learn the associations at all they should be at chance in all conditions.

Within each of the four pair conditions, half of the six deviant pairs were used for *Normal* trials and the other half for *Reversed* trials, resulting in 18 different trials per pair modality. Each trial appeared twice during the test, the order of the pairs within a trial and of the trials within the test was randomized. In total, there were 144 trials.

After being shown both pairs in a trial, participants were asked to select, using the left and right arrow keys, which one was more familiar to them, regarding what they had just seen during the previous exposure phase. After choosing a pair, participants were asked to report why they had chosen that one by selecting one of the following options (which were translated into Spanish) by pressing either the ‘Z’, ‘X’ or ‘C’ key: “I remembered the sequence”, “The sequence was more familiar to me” or “I guessed”. These answers correspond to the confidence conditions *Remember*, *Know* and *Guess*, respectively. We considered the *Remember* and *Know* conditions as indicators of explicit knowledge being accessed either through recall or familiarity, respectively, and the *Guess* condition as indicator of implicit knowledge.

The test was self-paced: participants had to press a key before each pair appeared and had no time limit to answer the questions, but they were encouraged to answer as fast as possible. Instructions were given to the participants in Spanish both printed on the screen and orally by the experimenter. We showed the participants a diagram of a trial while giving the instructions and made sure they had understood their task by asking them to repeat it.

3.1.4 Analysis and results

First of all, in order to know whether participants were able to recognize the standard pairs during the test phase we compared their performance against chance level in the different conditions. For the *Normal* trials, we ran four two-tailed one-sample *t*-tests on the participants mean test performances against chance level (50%). We used an alpha level of .05 as significance criterion for all statistical tests. In this analysis, because we performed multiple comparisons, we applied a Bonferroni correction and the alpha was corrected at .0125. Our results indicated that test performance was significantly above chance level on the *VV* $(M = 59\%, t(14) = 2.9, p = .012)$ and *AA* $(M = 61\%, t(14) = 3.02, p = .009)$ conditions; but not on the *VA* $(M = 54\%, t(14) = 1.1, p = .291)$ or *AV* $(M = 55\%, t(14) = 1.6, p = .131)$.

Next we performed the same analysis on the *Reversed* trials’ performance. We found no significant difference from chance level in any condition. The full results of the *t*-tests for both *Normal* and *Reversed*

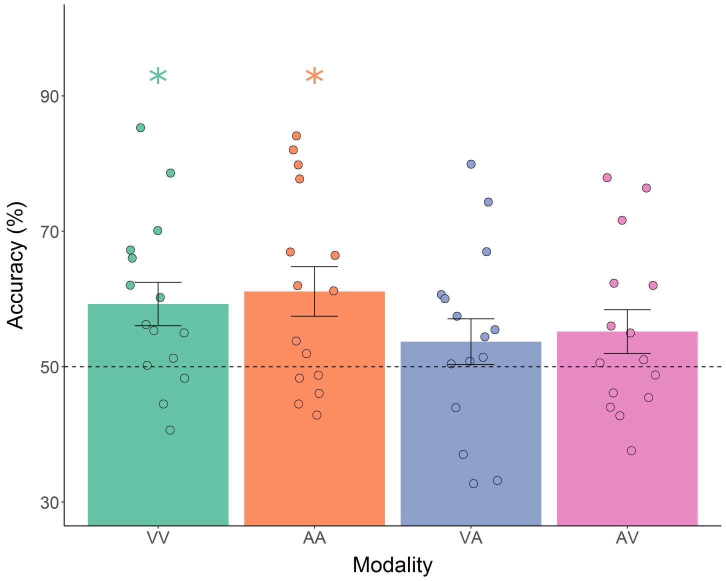

Figure 3.6: Mean accuracies and standard errors for familiar pair identification in *Normal* test trials during Experiment 1, separated by modality.

Table 3.1: Results of t-tests performed on the accuracies of *Normal* and *Reversed* trials against chance level.

Trial type	Pair modality	Mean	t	p
Normal trials	*VV*	59%	2.9	0.012*
	AA	61%	3.02	0.009*
	VA	54%	1.1	0.291
	AV	55%	1.61	0.131
Reversed trials	*VV*	56%	2.12	0.052
	AA	56%	1.97	0.069
	VA	53%	0.86	0.403
	AV	50%	-0.09	0.928

trials are shown in table 3.1.

Table 3.2: Results of χ^2 tests performed on the accuracies of *Normal* trials against chance level, aggregated by pair modality and reported confidence level.

Pair modality	Reported confidence	Sample size	Mean	χ^2	p
VV	“Remember”	68	66%	7.12	0.008
	“Know”	123	60%	5.08	0.024
	“Guess”	79	51%	0.11	0.736
AA	“Remember”	104	63%	7.54	0.006
	“Know”	126	61%	6.22	0.013
	“Guess”	40	55%	0.40	0.527
VA	“Remember”	75	48%	0.12	0.729
	“Know”	115	57%	2.51	0.113
	“Guess”	80	54%	0.45	0.502
AV	“Remember”	87	47%	0.29	0.592
	“Know”	116	57%	2.21	0.137
	“Guess”	67	63%	4.31	0.038

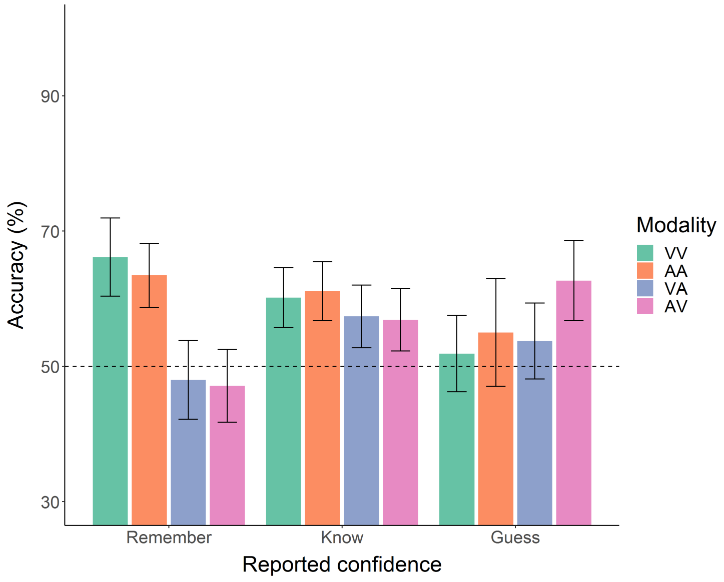

Figure 3.7: Mean accuracies and standard errors for familiar pair identification in *Normal* test trials during Experiment 1, separated by pair modality and reported confidence level. The standard error of each condition is directly proportional to the sample size of the condition.

To look at possible interactions between pair modalities and confidence ratings, we aggregated the test trials according to pair modality and reported confidence level and performed chi-square tests of independence for each condition combination. Results of all chi-square tests are presented in the following table. In this analysis the alpha was corrected at .0042 after applying Bonferroni correction. Unimodal trials (*VV* and *AA*) resulted in high performance when explicitly recognized, either by recall (*Remember*) or familiarity (*Know*), but not when guessed. When implicitly recognized (*Guess*), *AV* trials had the highest performance, $\chi^2(1, N = 67) = 4.31, p = .038$. However, none of the conditions had a performance significantly different from chance level according to our multiple comparisons correction criterion. The full results are shown in table 3.2.

3.1.5 Discussion

In this experiment we tested SL of both unimodal and crossmodal patterns. As we expected, participants were able to learn the unimodal pairs. Participants were not able to learn the crossmodal ones, suggesting that the SL mechanism is highly constrained within the different modalities. However this conclusion is not yet clear, since the lack of positive results in the cross-modal conditions could be attributed to other factors, such as the lack of power of the experiment due to the small sample size, the difficulty of the attentional task that may have left few cognitive resources for the SL mechanisms or the fact that we used an explicit measure of learning to measure knowledge that is many times implicit. Solving this last problem is what motivated us to perform the follow up EEG experiment, where we would be able to better detect implicit knowledge by looking at neurophysiological measures of learning.

When separating the results by confidence rating, we found in the unimodal conditions the trend that one would expect: participants performed better the more confident they were on their answers. In the crossmodal conditions, on the contrary, the opposite trend seems to appear, particularly in the AV condition, where accuracy was at its highest when participants reported just having guessed. None of this

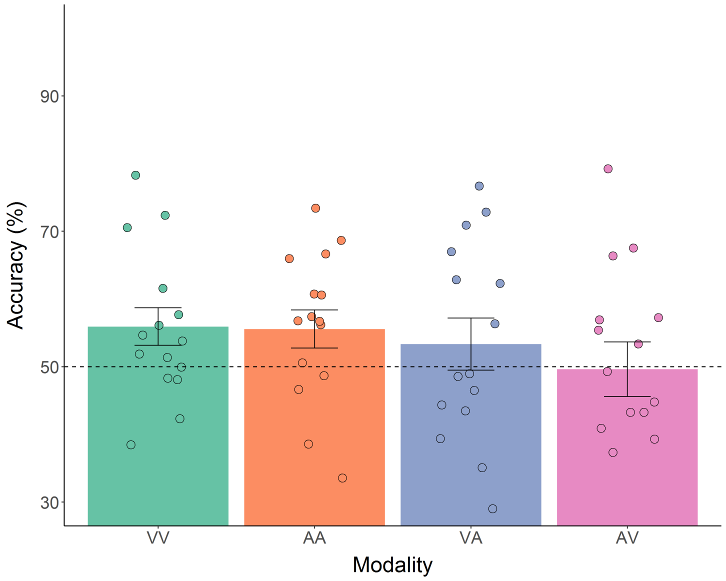

Figure 3.8: Mean accuracies and standard errors for familiar pair identification in *Reversed* test trials during Experiment 1, separated by modality.

was statistically significant, so these tendencies could be just noise due to the small sample size, but they also further suggested that there may be some implicit learning happening that we were missing due to the nature of the test, since explicit memory recall processes could interfere with implicit knowledge.

We found no significant differences from chance level in the performances of *Reversed* trials, so participants did not learn the pair associations irrespective of order.

A drawback of the experiment was the difficulty of the attentional task, that made us remove many participants (7) from the experiment because they failed to pass the oddball detection threshold that we set as a requisite. Upon examination, this was mostly due to the auditory oddballs and particularly to two auditory stimuli, for which the standard and oddball versions were not easily distinguishable. For the follow up experiment we replaced these two auditory stimuli by another two that where more easily distinguishable from their oddball versions and put more effort into making sure that participants were able to correctly perform the oddball detection task before the start the actual experiment.

3.2 Experiment 2: EEG experiment

3.2.1 Participants

Three naive adults (mean age of 23.3 years, all female) participated in the experiment in return for remuneration. Before the study, participants provided written informed consent. They reported no hearing problems and normal or corrected-to-normal visual acuity. One of the participants was removed from the analysis due to too many noisy channels making channel interpolation impossible.

3.2.2 Materials

For this experiment we used the same stimuli as in the previous experiment (except the two auditory stimuli that were replaced) and the pairs were created in the same manner.

The sequence orders of the stimuli pairs, oddballs and deviants within each of the subphases of the exposure phase (described below in the Procedure section) were created using Mix, a tool for generating pseudorandom sequences using constraint programming [25].

The timings for the presentation of the stimuli were the same as in Experiment 1 and a short break separated each phase.

3.2.3 Procedure

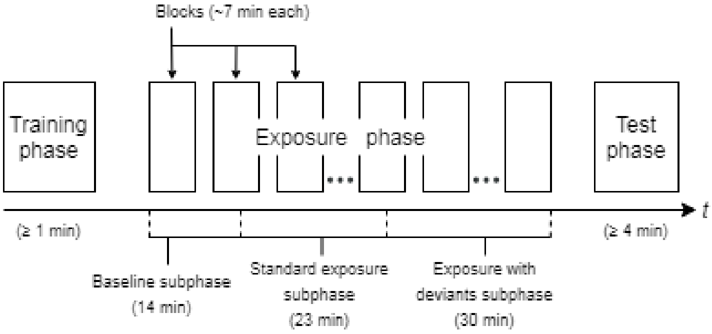

Figure 3.9: Diagram of the phases and subphases of Experiment 2

The experimental procedure was similar to the previous experiment, but with the following differences.

Some changes were made to the training phase of the previous experiment in order to better ensure that participants had understood the task and were able to reliable discriminate between oddball and standard stimuli. Each stimulus was presented in a random order four times, 50% of which as an oddball. During the training phase participants received online feedback about their performance with the appearance of a coloured ‘X’ of the same size as the fixation cross centered on the screen. The X is coloured blue whenever the participant misses an oddball stimulus (false negative) and it is coloured red whenever the participant incorrectly identifies a standard stimulus as an oddball (false positive). During this phase we considered an oddball stimulus as detected whenever a keypress occured no later than 750 ms after the stimulus’ onset. False negatives and positives were tracked in order to compute the participants F_1 score for each block separately. It was calculated as follows:

$$F_1 = \frac{2 \cdot \text{true positives}}{2 \cdot \text{true positives} + \text{false negatives} + \text{false positives}} \tag{3.1}$$

This score was shown to participants at the end of the phase. This training phase was repeated as long as participants F_1 score was below 80%.

The exposure phase consisted of three subphases: baseline, normal exposure and exposure with deviants.

In the baseline subphase all the stimuli were presented in random order. This subphase was constrained so that transitional probabilities between all different stimuli were equal, no stimulus repetition occurred and at least two standard stimuli appeared between any oddball stimuli. Each stimulus was presented during the baseline subphase 46 times, 2 of which as an oddball. Without taking into account time between blocks, this subphase had a duration of approximately 14 minutes.

The normal exposure subphase was like the exposure phase of Experiment 1. This phase was implemented in order to allow participant to learn the transitional probabilities. It was constrained so that transitional probabilities between all different pairs were equal, no pair repetition occurred and at least two pairs without oddballs appeared between any oddball stimuli. Each pair was presented 77 times during the normal exposure subphase. Each stimulus appeared twice as an oddball. Without taking into account time between blocks, this subphase had a duration of approximately 23 minutes.

During the exposure with deviants phase, some pairs had their trailing stimuli replaced by trailing stimuli of other pairs, in order to include violations of the already introduced statistical regularities. Two types of deviants were included: modality-coherent and modality-incoherent, which differed on whether the deviant stimulus is of the same modality as the replaced original trailing stimulus or not. The constraints used were the same as in the normal exposure subphase and the same maximum distance constraints were applied to deviant stimuli as to oddball stimuli. Each pair was presented 99 times during the exposure with deviants subphase, 10 of which included a modality-coherent deviant and other 10 included a modality-coherent deviant. Without taking into account time between blocks, this subphase had a duration of approximately 30 minutes.

During the exposure phase, participants were given no information about the statistical structure behind the stimuli or of the different subphases, which were concatenated seamlessly. The phase was divided into 10 blocks, each of them lasting for approximately 6 and a half minutes. Between blocks, participants were given the chance to rest and drink some water or juice and then resume the experiment by pressing a key whenever they wanted. After each block participants were shown their F1 score for that block. Here, we considered an oddball stimulus as detected whenever a keypress occured no later than 1.5 seconds after the stimulus' onset, thus giving participants some more leeway than during the training phase.

We included a test phase like the one of Experiment 1 in order to contrast behavioural with neurophysiological types of measure. Given the longer duration of the exposure phase during this experiment, in order to reduce the testing time and minimize participant fatigue this time we did not include *Reversed* trials.

3.2.4 EEG Recording and pre-processing

Electroencephalograms were recorded during the exposure and test phases using the BrainAmp DC (Brain Products) acquisition system and tin electrodes located at 29 standard positions (Fp1/2, Fz, F7/8, F3/4, FCz, Fc1/2, Fc5/6, Cz, C3/4, T3/4, Cp1/2, Cp5/6, Pz, P3/4, T5/6, PO1/2, Oz) mounted in an elastic cap. Electrode impedances were kept below 5 kΩ. The electrophysiological signals were digitized at a sampling rate of 250 Hz and filtered with an online high pass of 0.016 Hz. Eye movements were monitored with an electrode at the infraorbital ridge of the right eye. The ground electrode was located in the AFC position. The reference electrode was located at the side of the right eye. Scalp electrodes were later re-referenced offline to the mean activity of electrodes located at the left and right mastoids.

Electroencephalograms were recorded during the exposure (including all its subphases) and test phases, through only the signal within blocks in the exposure phase was pre-processed and analyzed. All preprocessing was done using the FieldTrip Toolbox for Matlab [26].

The recordings were subjected to a 0.1 Hz high-pass filter. Channels that exhibited many artifacts were interpolated using the distance-weighted average of their neighbouring channels. Channel neighbours were calculated using Delaunay triangulation.

The recordings were cut into epochs time-locked to the stimuli onsets that included the stimulus presentation period as well as the previous inter-stimulus-interval. 20 ms were removed both from the beginning and the end of the epochs in order to avoid occasional overlaps between epochs and the rejection of epochs because of the detection of artifacts right at their borders. This resulted in epochs with 230 ms pre-stimulus and 480 ms post-stimulus periods. Epochs corresponding to oddball stimuli or whose onset occurred up to 500 ms before or 750 ms after a keypress were removed from the analysis to avoid any motor preparation or motor execution interferences.

We identified (but did not reject at this point) blink artifacts automatically by thresholding the z-normalized Hilbert transform of the eye channel's signal. The threshold was chosen by visual inspection. In one participant the eye channel's signal was very low, so we transformed all other scalp channels instead and thresholded their average. Next we visually inspected the variability of the epochs in all channels both with and without using a 100 Hz high-pass filter to reject epochs containing muscle and jump artifacts respectively.

The epoched data of the scalp channels was then subject to an infomax-based ICA decomposition. The resulting components were visually inspected by looking at their topographies, their average activation across all epochs, their epoch variability during the whole experiment and their correspondence with the previously identified blinks. Eye-blink components were identified and rejected. Other artifact components were not rejected since they occurred in a small number of epochs, favouring the removal of such epochs. After the remaining components were backprojected to the original channel levels, we used the previously identified eye-blinks to remove epochs in which such a blink overlapped with the stimulus onset.

Finally, the epochs were passed through a 30Hz low-pass filter and their pre-stimulus signal was used for baseline correction.

3.2.5 Analysis

ERP Analysis

We computed average ERPs separately for each pair modality (*AA*, *VV*, *AV* and *VA*) and each of the following conditions: *Control*, *Predictable*, *Deviant stimulus* and *Deviant modality*. Within each pair modality condition (each having a leading and a trailing modality) we always compared stimuli of the trailing modality that were preceded by the leading modality. So for each pair modality, the *Control* condition consisted of stimuli taken from the baseline subphase whose modality was that of the condition's trailing modality and that were preceded by stimuli of the leading modality. These are stimuli that were not subject to any statistical regularities but shared the same modality switching effect as the stimuli they were going to be compared against. The *Predictable* condition consisted of the trailing stimuli of standard pairs that appeared during the exposure with deviants subphase, once the statistical learning mechanisms have supposedly taken place. The *Deviant item* condition consisted of the trailing stimuli of modality-coherent deviant pairs that appeared during the exposure with deviants subphase. The *Deviant modality* condition consisted of the trailing stimuli of modality-incoherent deviant pairs that appeared during the exposure with deviants subphase. Note that in the *Deviant modality* condition, given the change in modality of the trailing stimulus and in order to make meaningful comparisons between ERPs of the same modality, we compared effects of crossmodal learning against unimodal ERPs and vice versa. For example, a standard *AV* pair turns into an *AA* when appearing as a modality-incoherent deviant, so in order to notice effects of learning we compared their ERP against the *Control* ERP of the *AA* condition.

For each pair modality we visually compared the *Predictable*, *Deviant stimulus* and *Deviant modality* conditions against the *Control* condition. In the Predictable condition we looked for expectation suppression effects in ERP components, such as a reduction in the P3a component, which has been associated with stimulus novelty. In the Deviant stimulus and Deviant modality conditions we looked for surprise enhancement effects in ERP components, such as an increase in the P300 or N400 components, which have been associated with unexpected stimuli.

To plot the ERPs, we averaged the activity of the 4 channels with bigger variance in their grand average ERPs selected for each modality (visual or auditory) .

Neural decoding analysis

Next, we performed a neural decoding analysis using the temporal generalization method. The objective of this analysis was to get another neurological measure of learning to compare unimodal and cross-modal learning and to take a deeper look into the internal representations of the stimuli beyond what can be seen in an ERP. An important assumption of this analysis is that the spatial activation patterns of stimulus preparation are similar to the ones that would originate from processing the stimulus when no predictive mechanisms are involved [14]. For example, the pre-stimulus activation may share activation patterns with the post-stimulus activations of a family of similar stimuli, but they may not be precise enough to discern precisely which one of those stimuli was the one predicted. Another difficulty in making a comparison between pre- and post-stimulus activations are their differences in temporal dynamics: the post-stimulus activation varies greatly in time, so it is not trivial to make comparisons at which timepoint or time window we should look or, in general, how to deal with the temporal dimension.

Taking this into account, we decided to identify stimulus preparation by looking at the capacity of pre-stimulus activation to discriminate the modality of the stimulus. The activation networks of distinct modalities are different enough that we would expect to easily be able to discriminate their corresponding preparation activations.

To achieve this, we took a neural decoding approach using shrinkage LDA with temporal generalization. This analysis was performed separately for each participant and their results were later averaged. The data used, as in the previous analysis, were epochs that were composed of 230 ms of pre-stimulus activity and 480 ms of post-stimulus activity. We will refer to these epochs as trials. Decoders were trained and tested on the channel activity at single timepoints of the trials, so each sample consisted of 29 features, one per each channel.

Using temporal generalization, we were able to avoid the problem of the temporal differences between the pre- and post-stimulus activations, since it compares all timepoints with each other.

The decoders were trained to discriminate between the auditory and visual modalities on all the trials from the baseline subphase, since these trials were unaffected by any predictive mechanisms. The optimal γ value was calculated based on the average between the within-class covariance matrixes of auditory and visual data.

We balanced the training data in order to avoid biases towards a modality or towards a modality switch effect. The data were separated into four bins, depending on their stimulus modality and the modality of their preceding stimulus. To select which trials include in the training, we randomly sampled trials from each bin an equal number of times, matching the size of the smaller bin.

One decoder was built for each timepoint in an epoch, resulting in 179 decoders. These decoders were then applied to stimuli from the exposure with deviants subphase, separating between the unimodal and cross-modal conditions. Two conditions were considered: *Predictable* and *Deviant modality*. In the *Predictable* condition we included the trailing stimuli of standard pairs. In the *Deviant modality* condition we included the trailing stimuli of modality-incoherent deviant pairs. We considered the precise limit between the pre- and post-stimulus periods to be not the stimulus onset time, but the moment in which brain activity starts to hold information with discriminative power about the stimulus, thus ignoring the uninformative earliest perceptual processing stages. To pinpoint this moment we visualized the temporal generalization matrix of the baseline data. To get this matrix and avoid circular analysis, another batch of decoders was trained and tested on the baseline data (the same data used to train the decoders of the other conditions) by the means of 10-fold cross-validation. From the resulting temporal generalization matrix we could visually identify the limit between timepoints containing no information about stimulus modality (considered as the pre-stimulus period) and timepoints containing information about the stimulus modality (the post-stimulus period). Then we identified the area where markers of stimulus preparation were expected to appear. This area is the one corresponding to pre-stimulus timepoints being decoded by decoders trained on post-stimulus timepoints. We will refer to this area as the stimulus preparation area; it has been marked with a box in the temporal generalization visualizations.

In order to visualize the *Predictable* and *Deviant modality* conditions and avoid showing effects of modality switch, we created both an unimodal and a cross-modal *Control* condition, similar to the one used in the ERP analysis. However, instead of using stimuli from the baseline subphase (since these stimuli had been used for training and we wanted to use the same decoders as for the other conditions), we used leading stimuli of pairs of the exposure with deviants subphase that were either preceded by a stimulus of their same modality (to use as control and contrast for the unimodal conditions) or by a stimulus of the other other modality (to use as control and contrast for the cross-modal conditions).

By contrasting the temporal generalization matrix of the *Predictable* condition against the corresponding *Control* condition, a positive difference in accuracy in the stimulus preparation area would indicate that information regarding the stimulus modality was present during the pre-stimulus period, thus suggesting the presence of stimulus preparation.

The *Deviant Modality* temporal generalization matrix was contrasted against the *Control* matrix of the corresponding modality switch effect. In these contrasts we expected the opposite effect. A negative difference in accuracy in the stimulus preparation area would indicate information regarding the predicted stimulus modality (which is different to the modality of the presented stimulus), thus suggesting the presence of stimulus preparation. We included this condition because we expected the switch from predicting one modality to processing the other modality to be easily noticeable.

3.2.6 Results

Behavioural results

We performed no statistical analysis on these results and did not investigate the relationships between performance and reported confidence due to the small number of participants. Looking at the average performance, no pair modality seemed to have been reliably learnt. Participants did not perform better on

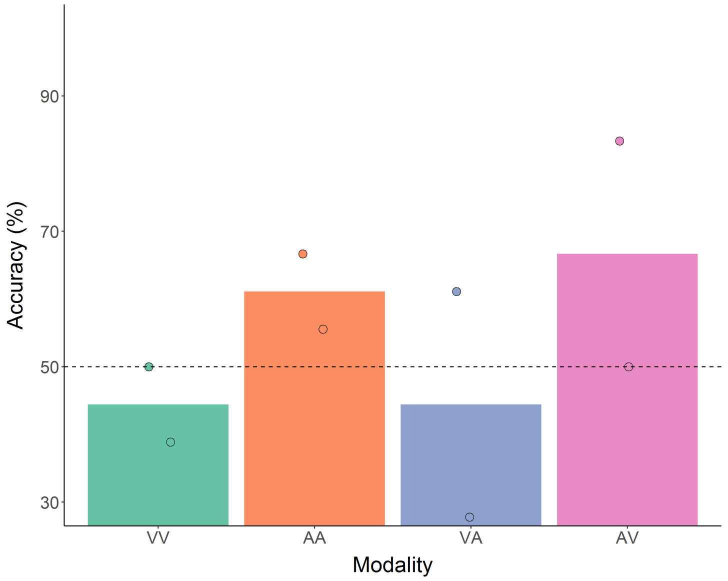

Figure 3.10: Mean accuracies for familiar pair identification in test trials during Experiment 2, separated by modality.

unimodal pairs than on multimodal pairs. They did perform better in conditions in which auditory stimuli worked as predictors (*AA* and *AV*) than on the conditions were visual stimuli were the predictors. This effect was due to one of the participants, while the other participant analyzed performed near chance level.

ERP results

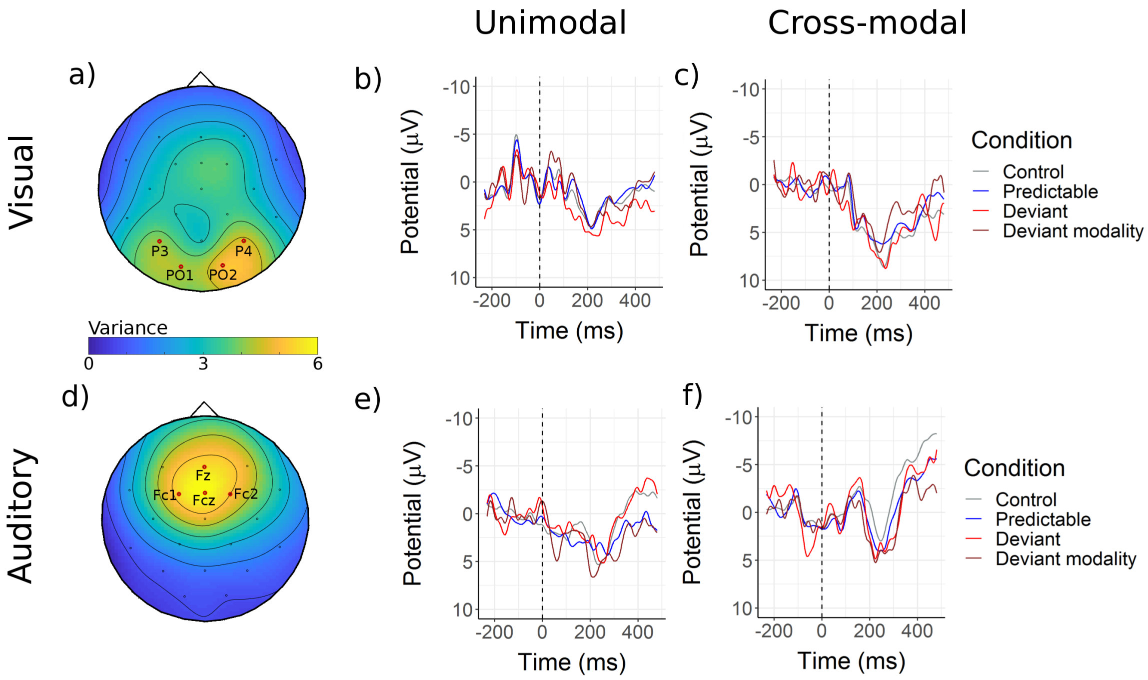

Figure 3.11: Results of the ERP analysis. a) and d) show the variance topography used to select which electrodes to plot.

We performed no statistical analysis given the small participant sample size. In each *Control* and *Predictable* conditions, around 200 epochs per subject were used to compute the average ERPs. In each of the *Deviant* conditions, only around 30 epochs were used to compute the average ERP. The small sample size for these last conditions means that their ERPs had a low signal-to-noise ratio and one should be careful when drawing conclusions from them. We visually inspected the ERPs to look for qualitative differences in the ERPs, which given a bigger sample size could result in true effects.

We noticed reductions of the N400 component with respect to the *Control* conditions in the *VV Deviant stimulus*, *AA Predictable* and *AA Deviant modality* (which corresponds to effects of learning *AV* pairs) conditions. The reduction in the *AA Predictable* N400 component could be due to expectation suppression, but we have no interpretation for the other N400 reductions.

In the *AV Predictable* condition, there seems to be a reduction in the P200 component, or an early P3a component; both components have been related to unexpected or novel stimuli, so their reduction could be due to expectation suppression. This is the clearest difference in components that we found and it agrees with the behavioural results, where *AV* was the pair modality that entailed a higher performance.

Some very early components also seem to appear in the unimodal *Deviant modality* conditions: an increase in the N100 component in the *VV Deviant modality* condition (which corresponds to an effect in learning *VA* pairs) and an increase in the P50 component in the *AA Deviant modality* condition. The N100 component has been associated to unexpected stimuli. P50 is an auditory component that is known to be reduced after stimulus repetition and has also been linked to attention. This component being higher than in the other conditions could be explained by participants expecting to see a visual stimulus (since this condition is associated with *AV* learning) and hence suddenly increasing attention and removing the P50 reduction effect that may be taking place in the other conditions. However we can not be sure about the validity of these effects due to the low signal-to-noise ratio in the *Deviant* conditions.

Neural decoding results

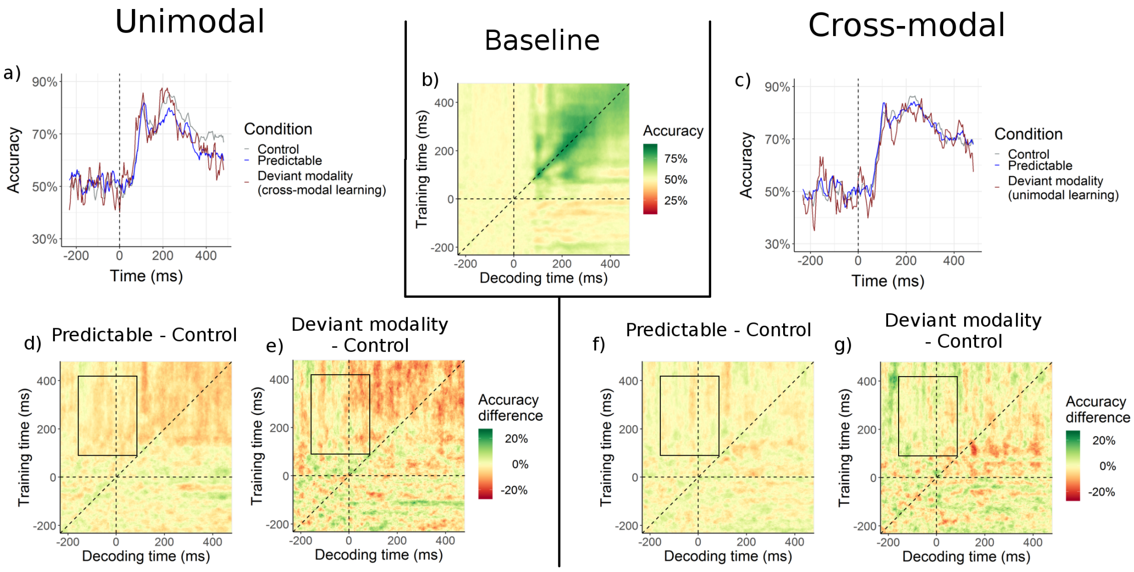

Figure 3.12: Results of the neural decoding analysis.

We did not notice any of the accuracy differences that we expected to see in the stimulus preparation area of the contrasts (*d*, *e*, *f* or *g*). A negative difference appears in stimulus preparation area of the *Unimodal deviant modality* contrast, but this negativity does not switch polarity or disappear once information of the stimulus appears along the *decoding time* axis (which is what we would expect to happen id this were stimulus preparation of a stimulus of a different modality), but it steadily continues during the rest of the epoch.

In fact, this negativity also appears in a weaker form in the *Unimodal predictable* condition and can be also noticed in the curves of the diagonal in *a*. We currently have no hypothesis as to what the cause of this difference in accuracy might be.

In the Baseline matrix, we can see a symmetric temporal generalization effect between the periods between 80-120 ms and 200-250 ms.

3.2.7 Discussion

The results of the behavioural test, although not significant due to the small sample size, were very different to the ones obtained in Experiment 1. This could be due to an important strategic component involved in the task, resulting in a higher within-groups than between-groups variability in the test results that makes impossible to infer the real group variability from a small sample size. This strategic variability may have been increased by the lack of Reversed trials in this test, so we could not be sure of whether the measured performances were a product of real learning that occurred during the exposure phase or of inferential strategies that had taken place during the test. These inferential strategies could also explain below chance performances, such as the one observed in one participant in the *VA* condition, in case participants had started their inferential processes from an incorrect assumption.

In the ERP analysis, we saw some evidence suggesting learning of *AA*, *VA* and *AV* pairs. Only the *VV* condition showed no signs of learning, which coincides with the behavioural results, since the *VV* condition was the only one where both participants performed at or below chance level. However, we also saw some effects that could not be attributed to learning and were of similar size to the ones explainable by learning, namely the reduction of the N400 component in the *VV Deviant stimulus* and *AA Deviant modality* conditions. So, variations in the N400 component may be due to noise or other processes different from learning. The N400 component may not be a good measure of learning for this paradigm, since according to some previous research [27], N400 surprise enhancement only takes place when semantic mismatch happens. Thus it would require the stimuli to be semantically meaningful, precisely what we were trying to avoid in our paradigm.

The clearest effect we noticed was the P200 or P3a decrease in the *AV Predictable* conditions. Differences in the 200 - 300 ms have been associated to stimulus novelty (and thus their reduction to expectancy) without the need of semantic associations [28][29]. This effect agrees with the results of the behavioural test, where the best performance was in the *AV* modality, therefore strengthening the conclusion that *AV* learning really took place.

Unfortunately the neural decoding analysis did not show any of the signs of stimulus preparation that we expected to notice. This may have been due to no learning, and thus no stimulus preparation, taking place; or only learning a very limited number of pairs, so stimulus preparation would appear in too few trials to be noticeable above the noise level that is present in the contrasts; or due to stimulus preparation encoding the modality in a different way than the actual, so that its decoding can not be generalized in this way.

Chapter 4

Final discussion and conclusions

This chapter summarily discusses the thesis findings, which allow us to reach a number of conclusions. Importantly, it also outlines a few potential future lines of research stemming from the thesis work.

In Experiment 1, we found that participants were able to explicitly learn unimodal but not cross-modal patterns of non-meaningful auditory or visual stimuli while performing an orthogonal attentional task. However, implicit learning of cross-modal *AV* pairs might have also occurred.

In Experiment 2, we found some neurophysiological evidence indicating the presence of cross-modal *AV* learning. However, we cannot extract general conclusions from these particular findings, given the small sample size and the high individual variability that seems to affect the task.

In general, the experimental thesis findings support our second hypothesis, namely that SL is a domain-specific system with high constraints between modalities, thus agreeing with much of existing previous research.

A surprising result that did not agree with the second hypothesis that we stated at the onset was the learning of *AV* pairs. We hypothesised that in the case of SL being domain-specific, any cross-modal learning would be more explicit than the unimodal learning, since it would be due to higher-level cognitive processes that required explicit recognition of the stimuli. However, the results of Experiment 1, although not significant, seemed to suggest the opposite to be true.

The reality of the SL system may be much more complex than discerning whether it is domain-general or domain-specific and its differences from other types of learning is not completely clear[30] [31]. Even though similar computational processes seem to take place in each modality, they also have very distinct constraints (temporal, contextual, attentional, etc.) and integrate with the rest of modalities in different ways. These constraints and cross-modal relationships may have evolved *ad hoc* in order to solve the different problems faced in our environment. For example, the ability to learn dependencies from auditory to visual stimuli may have been crucial to our survival in order to predict the appearance of predators before having them on sight, but the opposite dependency (from visual to auditory) may not have been as useful.

Our experiment suffered from some drawbacks. The principal one being the small sample size, particularly in the second experiment, that did not allow us to extract meaningful conclusions in many analysis. Also, the attentional task included appeared to be very difficult to many participants and it is unclear whether it helped or hindered the SL processes. Attention seems to play a crucial role in SL [32][2], although its role is not completely understood and not agreed upon in the literature. In hindsight, we should have callibrated better the difficulty of the attentional task.

4.1 Conclusion

We wanted to discern whether the statistical learning system is domain-general or domain-specific. Our findings showed that it is much harder to learn cross-modal patterns than unimodal ones, revealing that there are big constraints between modalities and that SL is a primarily domain-specific cognitive mechanism.

4.2 Future work

There are several ways in which this research could be continued. Clearly more participants are needed for both experiments, particularly for Experiment 2, in order to get more conclusive results. Besides getting more participants, we could obtain more information about the learning processes going on without the need of many changes to the experimental paradigm by performing new analysis. For example tracking the neurophysiological responses to individual items to see which pairs were the ones being learnt, whether they correspond to the ones being correctly identified in the test phase and whether there are other confounding factors that favor the learning of some dependencies over others and that we did not control for. We could then investigate the evolution of the neurophysiological responses throughout the exposure, in order to obtain a learning curve. This curve could also be obtained by correctly implementing the oddball detection reaction time measure that we initially intended to use. Obtaining learning curves opens many possibilities, since we could notice other differences between unimodal and cross-modal learning that would be hidden otherwise and qualitatively compare the curves with the ones resulting from different computational models of learning.

A blocked experimental design, where cross-modal and unimodal are learnt separately would also be very interesting to conduct, since it would avoid the possible interference between different types of learning. There are also additional experimental factors that could be included into the experimental design, such as variations in the TP in order to add uncertainty or variations in the difficulty of the attentional task, to see how much an attentional task

Finally, it would be very interesting to further explore the possibilities of temporal generalization in neural decoding. It is a fairly new analysis method that can show us many things about the dynamics of cognitive processes.

Bibliography

[1] Jenny R. Saffran, Richard N. Aslin, and Elissa L. Newport. Statistical learning by 8-month-old infants. *Science*, 1996.

[2] Ana Franco, Vinciane Gaillard, Axel Cleeremans, and Arnaud Destrebecqz. Assessing segmentation processes by click detection: online measure of statistical learning, or simple interference? *Behavior Research Methods*, 47(4):1393–1403, 2015.

[3] Dilshat Abla, Kentaro Katahira, and Kazuo Okanoya. On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, 20(6):952–964, 2008.

[4] Arit Glicksohn and Asher Cohen. The role of cross-modal associations in statistical learning. *Psychonomic Bulletin and Review*, 20(6):1161–1169, 2013.

[5] Aaron D. Mitchel and Daniel J. Weiss. Learning Across Senses: Cross-Modal Effects in Multisensory Statistical Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, 37(5):1081–1091, 2011.

[6] Ivan P Pavlov. Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. *Oxford University Press*, 1927.

[7] John B. Watson and Rosalie Rayner. Conditioned emotional reactions. *Journal of Experimental Psychology*, 1920.

[8] Jun Ichiro Kawahara. Auditory-visual contextual cuing effect. *Perception and Psychophysics*, 69(8):1399–1408, 2007.

[9] Sarah C. Creel, Elissa L. Newport, and Richard N. Aslin. Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30(5):1119–1130, 2004.

[10] Joel L. Voss and Ken A. Paller. An electrophysiological signature of unconscious recognition memory. *Nature Neuroscience*, 12(3):349–355, 2009.

[11] Dilshat Abla and Kazuo Okanoya. Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research*, 64(2):185–190, 2009.

[12] Susheel Kumar, Peter Kaposvari, and Rufin Vogels. Encoding of Predictable and Unpredictable Stimuli by Inferior Temporal Cortical Neurons. *Journal of Cognitive Neuroscience*, 29(8):1445–1454, 8 2017.

[13] Peter Kaposvari, Susheel Kumar, and Rufin Vogels. Statistical Learning Signals in Macaque Inferior Temporal Cortex. *Cerebral Cortex*, 28(1):250–266, 2018.

[14] Peter Kok, Michel F. Failing, and Floris P. de Lange. Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, 2014.

[15] Pim Mostert, Peter Kok, and Floris P. De Lange. Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, 5(July):1–13, 2015.

[16] J. R. King and S. Dehaene. Characterizing the dynamics of mental representations: The temporal generalization method, 2014.

[17] Julie Onton and Scott Makeig. Information-based modeling of event-related brain dynamics. *Progress in brain research*, 2006.

[18] Scott Makeig, Stefan Debener, Julie Onton, and Arnaud Delorme. Mining event-related brain dynamics, 2004.

[19] Anthony J. Bell and Terrence J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 1995.

[20] Arnaud Delorme and Scott Makeig. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 2004.

[21] Jerome H. Friedman. Regularized discriminant analysis. *Journal of the American Statistical Association*, 1989.

[22] Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 2005.

[23] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus Robert Müller. Single-trial analysis and classification of ERP components - A tutorial. *NeuroImage*, 56(2):814–825, 2011.

[24] Scott Draves and Erik Reckase. The Fractal Flame Algorithm. *Unpublished manuscript*, 2008.

[25] Maarten Van Casteren and Matthew H. Davis. Mix, a program for pseudorandomization. *Behavior Research Methods*, 2006.

[26] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan Mathijs Schoffelen. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011.

[27] Marta Kutas and Steven A. Hillyard. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 1980.

[28] David Friedman, Yael M. Cycowicz, and Helen Gaeta. The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty, 2001.

[29] Robert F. Simons, Frances K. Graham, Mark A. Miles, and Xun Chen. On the relationship of P3a and the Novelty-P3. *Biological Psychology*, 2001.

[30] Todd M Gureckis and Bradley C Love. Behaviorism Reborn ? Statistical Learning as Simple Conditioning. *Cognition*, pages 335–340, 2007.

[31] Pierre Perruchet and Sebastien Pacton. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5):233–238, 2006.

[32] Nicholas B. Turk-Browne, Justin A. Jungé, and Brian J. Scholl. The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, 134(4):552–564, 2005.