Measuring and mapping brain response to congruently and incongruently placed objects in relation to the background

Nick Verhagen

Sunday 11th September, 2022

1 Abstract

Human beings are generally excellent at rapid object recognition, especially in isolation. However, objects are not viewed in isolation in the real world; instead, they are viewed within a context. This paper explores brain activity based on fMRI-measured blood-oxygen-level-dependent (BOLD) response using a task that is designed to test the ability of the brain to distinguish between congruent and incongruent objects within a scene. In this context, the quality of "congruency" describes the relationship between expected object size and the scene in which the object is placed. Analysis of the data obtained from this experiment showed a significant difference between brain activity when viewing an object in an incongruent or congruent context. Further, we can use support vector classifier (SVC) models to more sensitively predict whether a participant was looking at a congruent or incongruent object better than random chance. The findings of our analysis are consistent with current research in the field and offer a more detailed study of this neuroimaging finding. In our analysis, the fusiform face area was also found to be a strong predictor of congruency - a finding that is consistent with the most recent research in the field. Further, using searchlight analysis, we find voxels that discriminate on congruency in line with existing literature: the lateral occipital cortex and early visual cortex. However, areas that have received less attention are also strongly discriminant on congruency based on this analysis, namely the prefrontal cortex and caudate nucleus. Owing to these areas' high activity during expectation-violations, we can conclude that the lateral occipital cortex and the early visual cortex are important to determining size-context relationship violations. However, the prefrontal cortex, caudate nucleus, and potentially other areas are also highly active and show congruency discrimination. This suggests involvement of a large part of the brain during object recognition and expectation-violation, and not the compartmentalised view of this process as older literature suggests. Lastly, activation within the fusiform face area was found to be a strong indicator of congruency.

2 Introduction

Humans are very efficient at object recognition, and there is much research into the recognition of isolated objects. However, in the real world, objects are not viewed in isolation, but in varying contexts. Earlier work has shown that object recognition is influenced by context (Biederman et al., 1982) (Ganis & Kutas, 2003), so that we may better recognise a boat that is in water than a boat that is seemingly floating.

However, is there a noticeable change in brain response when looking at such an image, where the context does not match the object of focus? That is the goal of this analysis: Is there a significant difference between brain activity while viewing congruent objects and while viewing incongruent objects - and if so, in which direction is this difference? Secondarily, where in the brain does this difference occur, and why is this the case? To explore this, fMRI imaging is used, along with an experimental setup that will allow distinguishing between images representing different conditions of congruency.

There has been much research completed in the past regarding scene and object perception and the factors that influence it eg. (Biederman et al., 1982) (Cate et al., 2011) (Perini et al., 2020) (Grill-Spector et al., 2001) (Rolls, 2001). Specifically for this paper, we will be looking at the difference in brain activity between congruently and incongruently placed objects within a scene; congruent describes an object that matches the context it is within, while incongruent describes an object that does not. In simpler terms, in the paper by Biederman et al. (1982), congruency is described as 'familiar size'. Objects are expected to be a certain size compared to others, and it is congruent when the size matches what is expected, and incongruent when not. In a practical example, we expect a person in the distance to be a small size as it is projected on the retina, and a person that is close to be a larger size when projected on the retina. If this is not the case, we consider it incongruent.

Congruency as it is used in this thesis is a concept that lies within the concept of expectation-violations. Expectation, and priming, is a behaviour of the brain that occurs constantly, facilitates learning, and shapes context (Wang et al., 2004) (Ericsson et al., 1993) (Ericsson & Lehmann, 1996). Even from a very early age, for example, infants hold expectations for perceived object size and are surprised when an object is entirely hidden by an occluder that seems too narrow (Wang et al., 2004). Congruency in the way it is used in this thesis is closely related, as it concerns expectations of object size and distance relative to scene context.

Through years of research, different areas of the brain were identified to be important for object recognition (Khateb et al., 2002) (Grill-Spector et al., 2001) (McGugin et al., 2016). This is a very active field of research, and our understanding of it will likely shift over time. However, for the purpose of

this analysis, we will rely on the current understanding to guide which areas to focus on for the task outlined in this paper. Classically, the fusiform face area (FFA) was considered to be primarily involved in the recognition of faces. However, some literature suggests the FFA informs general object recognition as well (McGugin et al., 2016) (Gauthier et al., 1999). As such, the FFA is given additional attention within this paper in regards to distinguishing between congruent and incongruent images.

Other areas of the brain are often considered to contribute strongly to object recognition. For this paper, the early visual cortex (EVC) and the lateral occipital cortex (LOC) will be particularly important. Literature shows that the LOC strongly contributes to the perception of the size and shape of objects (Cate et al., 2011) (Grill-Spector et al., 2001). With the knowledge that congruency as it is defined in this paper hinges on object size context violations, the LOC is of particular interest. The EVC is also important, as it is found to be active in many stages of visual computation (Lee, 2003). Early models of visual processing considered every step to be quite encapsulated, with the EVC being primarily involved in the early stages (Marr, 1982). In either scenario, the EVC is of high importance for object recognition. Further, the EVC exhibits increased activity when primed to receive visual stimulus, prior to the arrival of the stimulus (Giesbrecht et al., 2006). The LOC also exhibits a higher performance in object recognition when preparatory activity is performed (Peelen & Kastner, 2011). These findings in existing literature cement the importance of these areas in object recognition, especially as it pertains to expectation. This relates them closely to the brain's processing of incongruency, being a visual expectation-violation of size and distance (Biederman et al., 1982).

Expectations and its violations are a broad concept, and where the definition of congruency in this case lies within the neurological, the concept applies to motor control (Grush, 2004) too, for example. There is a considerable amount of literature, including fMRI studies, that concerns itself with expectations and priming (De Lange et al., 2018). This is seen even outside the realm of purely object recognition - for example, expectation-violations in magic tricks, as in Danek et al. (2015). These studies also suggest brain regions that experience high activity during expectation-violations, such as the caudate nucleus (CN) and prefrontal cortex, which are normally not discussed in the context of object recognition. It is important as a consequence to consider the scope of this thesis as being entirely on visual object recognition and expectation-violations as a result of congruency during this task.

To answer the primary research question: "Is there a significant, detectable difference between brain activity while viewing congruent objects and while viewing incongruent objects - and if so, in which direction is this difference?", there will be various methods of analysis used on the data obtained from the experiment. Using increasingly sensitive methods, first, correlations and univariate contrasts are used to address this primary research question. Following

these, as a more sensitive method, a support vector classifier (SVC) model will be trained on all regions of interest separately to create models that can predict congruency based on brain activity as measured by fMRI.

Lastly, a full-brain searchlight analysis will be performed to answer the secondary question - "where in the brain does this difference occur, and why is this the case?" Searchlight analysis is a relatively new form of multivariate pattern analysis (Etzel et al., 2013). Quoting Etzel et al. (2013): "Searchlight analysis produces maps by measuring the information in small spherical subsets ("searchlights") centered on every voxel; the map value for each voxel thus derives from the information present in its searchlight, not the voxel individually." This method will be used to find clusters of voxels that seem to aid in distinguishing between congruent and incongruent stimuli, and whether these clusters are found within areas that are expected in comparison to the literature, eg. Grill-Spector et al. (2001) Lee (2003) McGugin et al. (2016). Several areas are of note according to existing literature, which is where clusters are expected to be found as a result of the searchlight analysis. Perini et al. (2020) finds the prefrontal cortex to be active following a searchlight analysis into haptic object size, and Danek et al. (2015) and Schiffer and Schubotz (2011) find that the caudate nucleus (CN) is highly active in expectation-violation tests. Further, Giesbrecht et al. (2006) and Peelen and Kastner (2011) find heightened activity in the LOC and EVC when primed for recognition tasks. This is reason to expect to find some clusters of voxels in these areas as well upon our own searchlight analysis, as we can expect a difference in response when processing stimuli differing in congruency.

3 Method

3.1 Task and procedure

This paper's experimental setup and the resulting data is based on a currently unreleased paper by Gayet et al., in which an fMRI machine is used to conduct the following experiment. Instructions are given verbally to the participants. During the experiment, participants are shown stimuli in the form of images for a total of 336 seconds per 'run'. A single session lasts roughly 1.5 hours, comprising of the explanation of the procedure, numerous "runs", and miscellaneous pauses. Participants were instructed to look straight ahead at a grey background, on which the visual stimuli is shown. Before any stimuli is shown in a run, participants are instructed to look at the grey background for 16 seconds. This is called a baseline fixation block. Afterwards, stimuli is shown for 150 ms, with 850 ms of "off" space, only showing the grey background, directly after. This is to reduce eye movement. If they were shown the same image twice, they were to press a button. This task was given to keep participants focused on the

stimuli and produce the best results possible, optimising the blood-oxygen-level-dependent (BOLD) response measured by fMRI. A set of 16 images is called a block, or mini-block, which follow a set of conditions that apply to all of images within the block. After a block of images has been shown, a new block is shown immediately after. The new block's conditions are random, but can not be the same as the block before. After 4 blocks have been shown, a baseline fixation block is shown for 16 seconds, containing only the grey background. This is also done at the end of a run.

3.2 Setup

As mentioned, the experiment is based on fMRI results. Within this section the specifics of the setup for the fMRI machine are outlined. The fMRI settings are the same as in Gayet and Peelen (2022), directly quoted below:

- "Participants view the stimuli through a mirror mounted on the head coil of the scanner"
- "The effective viewing distance (eyes-mirror + mirror-screen) approximated 1440mm."
- "Stimuli were presented on a 1024 x 768 EIKI LC – XL100 projector (60 Hz refresh rate), back-projected onto a projection screen (Macada DAP diffuse KBA) attached to the back of the scanner bore."
- Participants are given a button connected to the serial port of the computer handling the presentation.
- "fMRI data were acquired on a 3T Magnetom PrismaFit MR Scanner (Siemens AG, Healthcare Sector, Erlangen, Germany) using a 32-channel head coil."
- "A T2*-weighted gradient echo EPI sequence with 6x multiband acceleration factor was used for acquisition of functional data (TR 1 s, TE 34ms, flip angle 60, 2 mm isotropic voxels, 66 slices)."
- "A high-resolution T1-weighted anatomical scan was acquired at the start of each experimental session, using an MPRAGE sequence (TR 2.3 s, TE 3.03ms, flip angle: 8, 1 mm isotropic voxels, 192 sagittal slices, FOV 256 mm)."

3.3 Stimuli

The set of images that comprises the stimuli for the participants are the 'best' 64 scenes from an also yet unreleased study by Gayet et al., in which participants

were to distinguish between animacy (living objects) and inanimacy (nonliving objects). The 64 scenes used in the following fMRI study are those that produced the strongest difference in animacy discrimination between congruent and incongruent conditions from a total of 86 initial scenes. Within this thesis, animacy is not otherwise used to determine congruency.

All scenes comprise of an object that matches any of the 4 possible combinations of congruency and animacy, placed within a scene and degraded by pixelation to increase the potential influence of scene context, as can be seen in figure 1. Every object is placed in either a 'far away' or 'close' position, the object changing in size as projected on the retina to match this distance, creating the congruent condition. To create the incongruent condition of the same scene-object combination, the two distances are swapped. The object that is large as projected on the retina is in the 'far away' position, and the object that is small on the retina is in the 'close' position. The images within a block are random, though they all adhere to a given combination of congruency and animacy. Within a block, large and small objects are intermixed.

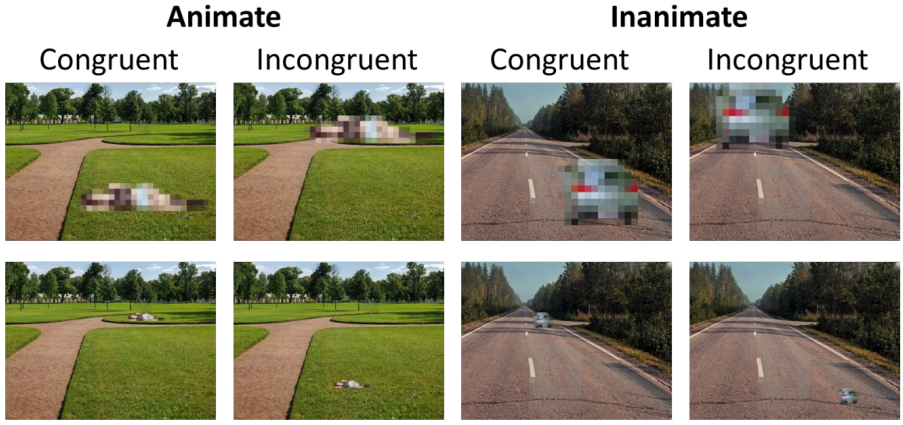

Figure 1: Example of the 8 different experimental conditions

Above is an example of stimuli shown during a run. As can be seen, congruency of the subject hinges on the subject not violating the expectation of its size and location in relation to the scene. In simpler terms, the subject of a stimulus is congruent when its size and location within the scene make it appear as an object that 'makes sense' to the viewer and matches their expectations. For example, the person depicted is incongruent when they appear unusually large or unusually small due to their absolute size and their position within the scene.

3.4 Data

For every second of a 336 second run, an fMRI scan was performed and stored in a .nii file after preprocessing. As such, every run of 336 seconds produces 336 files. Every scan contains 592,895 features, in the shape of 79*95*79, detailing BOLD response as a number for every 2mm voxel. The four participants, with 6 runs from each participant, were used as the data for this paper.

3.5 Analysis

Various methods of analysing the data were employed. Primarily, correlation analysis and univariate contrasts (comparison of mean activity within a ROI between the two conditions of interest) were used to explore if there was a significant difference between incongruent and congruent objects. The more sensitive SVC models were used to determine whether or not activation signals from specific regions of the brain can be used to predict if an object is congruent or incongruent.

First, every set of data belonging to a single person, that is, all 6 runs and the 336 scans per run, was loaded into separate Numpy arrays for ease of handling. Since we were provided with the data of 4 people total, this means there are 4 arrays. Besides initial data exploration, the majority of data analysis was completed on the first person. There are two main reasons for this decision. First, time constraints had to be considered, with some operations taking a considerable time even when just applied to the data of a single person. Secondly, the data differs quite strongly between subjects, and analyzing the four subjects as a group would not be informative. This will be further explored in the results and conclusions.

After loading in the data, before filtering, blocks were sorted by means of regex into congruent and incongruent. To account somewhat for the delay in hemodynamic response, there are an added 4 seconds onto the first time a stimuli for a specific condition is shown, continuing through to the end of the block. This roughly captured most of the peaks of the hemodynamic response. The 4 second extension also applies to the end of the block, which means it captures scans that belong to a different block. However, response to the new stimuli should be minimal due to the delay in BOLD response. The baseline fixation blocks were not included. Scans belonging to a block were grouped together, making for groups of 16 arrays. When considering all runs for a single person, 42 groups of 16 second long miniblocks which contain a brain scan per second were made, per condition. Congruent and incongruent had the same amount of scans every run, both ending up with 42 groups over all runs belonging to a single person.

As fMRI data does not correlate strongly between-subjects, as explained within the appendix, I made the decision to limit analysis to a single person. After making this decision, an analysis was performed comparing the means of miniblocks between congruent and incongruent, based on a single person, to optimise the similarity of data. To transform the data, the activity within a second was first averaged, creating an array in the shape of 42 * 16, 42 miniblocks and 16 seconds of averaged activities in a miniblock. This array was then averaged again on its second axis (axis = 1 in Python), which averaged the data on a per-miniblock basis. This creates a 1-dimensional array in the shape of 42, 42 averaged miniblocks. These steps were taken for both the congruent and incongruent data. Two raincloud plots were created using the newly created arrays to obtain a clear overview of the data and to determine differences between congruent and incongruent at a glance. This was done for 3 different regions of interest. For each of these, a Boolean index was created corresponding to the voxels within that ROI. This was then applied to the data, leaving out voxels that did not correspond to it. As described further in the introduction, the chosen regions of interest were the early visual cortex (EVC), lateral occipital cortex (LOC) and the fusiform face area (FFA), as these areas are likely to produce strong responses (Gayet & Peelen, 2022) (Khateb et al., 2002). While normally seen as an area of the brain that concerns itself mainly with face recognition, literature suggests the FFA is also linked to object recognition performance - it was included here as a result (Gauthier et al., 1999) (McGugin et al., 2016). The statistical significance of the findings from this initial analysis were examined in the next analysis where the differences are examined more closely.

To determine if there are systemic differences between congruent and incongruent miniblocks, the mean difference between congruent and incongruent miniblocks was determined and plotted for every ROI. The difference per miniblock is on a by-element basis - as there were 2 arrays with 42 elements, 42 differences were obtained by taking the difference between every same-number element. This is done in this manner to find differences between blocks that are as similar to one another as possible, minimising differences in activation over time due to other factors. Following this, error bars were created by determining the SEM (standard error of the mean) of the differences and plotting these on top of the means. Lastly, one-sample t-test against zero was performed for every ROI to determine the significance of the systemic differences.

To determine the amount of information present in the univariate form of the data, the next statistical test performed was a correlation analysis limited to a single person. This should inform whether it is expected to find positive results from training SVC models. If this analysis is statistically significant, then we can expect statistically significant findings for an SVC as well. Furthermore, the analysis will further reinforce the findings within the previously performed univariate contrast analysis. To perform the analysis the data was split into 4 groups total, 2 congruent and 2 incongruent. The data was shaped in such a way as to be per-second mean activation in the brain. This created two

arrays in the shape of (672,), one for congruent data and one for incongruent data. The two initial groups, congruent and incongruent, were split by sorting equal and unequal indices into two separate groups. This creates 4 separate arrays in the shape of (336,). Splitting the data in this way optimised similarity between the seconds and minimised comparing activation between runs. If this were not done, the variation in activation between runs could possibly account for a high amount of the correlation, or lack thereof. As before, data that is limited to the region of interest EVC is used. We expect to see highly correlated results, as voxel activity does not vary extremely from second to second (Friman et al., 2002). If the results correlate in the way we expect, there should be higher correlations between data that corresponds to the conditions of interest. So, if congruent values have a higher correlation with congruent values than with incongruent values, then we could expect to be able to train a model off of this with accuracy above that of random guessing. Whether there is a sufficiently large difference for this purpose is done by applying the formula: corr((congruent1-congruent2)+corr(incongruent1-incongruent2)) - corr((congruent1-incongruent2)+corr(congruent2-incongruent1)) over random halves of the earlier mentioned data, 10000 times. Statistical significance was determined by whether the random groups can beat the difference found in the real data, to inform a p-value.

Next, to examine with a more sensitive analysis how well the heightened activity when viewing congruent stimuli can be used to distinguish between congruent and incongruent images, an SVC algorithm was used on the data belonging to person 1. This included all their runs. EVC, LOC and FFA data were considered separately. This is true for calculating overall accuracy in repeated tests, and for the grid search as well. The data was further prepared by reshaping it into the right form so that is usable by the SVC algorithm; specifically, labeling each second with the appropriate condition and reshaping the data into a 2-D array. This means it is an array that contains every relevant second, and in each array of each second, there is an array corresponding to the amount of voxels and its values. The voxels were normalised by z-score. While more computationally intense to consider every second, rather than averaging or normalising over full blocks, the highest accuracy was obtained this way. Hyperparameter tuning was applied by use of grid search, with cross-validation. In cross-validation, 25% of the dataset was used as a test set. Below, in table 1, is the grid of possible parameter values. Note that for eventual testing, all of the data was used as both training and testing by making random groups for each time an SVC model is created. This was not done for grid search, primarily due to time restraints. 25% of the data was used as a test set, and this same data was used for the entire grid search.

Kernel	C	Gamma
Linear	0.01, 0.05, 0.1, 1, 10, 100, 1000	10, 1, 1e-1, 1e-2, 1e-3, 1e-4
RBF	0.01, 0.05, 0.1, 1, 10, 100, 1000	10, 1, 1e-1, 1e-2, 1e-3, 1e-4

Table 1: Table containing all tested hyperparameters through hyperparamater tuning

Following this grid search with cross-validation, two different sets of parameters came out as possibilities for highest precision, as follows: RBF kernel with 10 C and 0.001 gamma, and linear kernel with 0.01 C and 10 gamma. Of these, the RBF kernel was chosen, as this produced higher accuracy when running the grid search a multitude of times, and when further acquiring results.

For calculating the accuracies with higher confidence, the full data was randomly split 40 times each for all 3 versions, then an SVC model was trained on the split data, and an accuracy value calculated based on that. 40 repeats were completed to acquire arrays of 40 accuracies each. For each accuracy, a new split was used and had a new model trained on it, acquiring accuracies that are independent of one another. Every array was then one-sample t-tested against an array containing the expected accuracies of random guessing. The expected accuracy is 0.5, as there are an equal number of congruent and incongruent samples.

Finally, a searchlight analysis was performed. This was done to get a detailed, multivariate, localized look at the voxels and areas of the brain that can distinguish between activity evoked by viewing size-congruent and incongruent objects. This was then compared against our understanding of important areas for object recognition and expectation-violations, informed by papers such as Grill-Spector et al. (2001), Wang et al. (2004), Perini et al. (2020), and others.

To do this, the 3D Nifti files of a single run from a single person, congruent and incongruent but excluding the baseline miniblocks, were transformed into a single 4D Nifti file (containing the number of the sample on the 4th dimension). This was done using Matlab, specifically SPM and its batch functionality. From this 4D Nifti file, a mask containing all of the brain's voxels was created. This was done by using the image function from the Python package, Nilearn. Specifically, voxels that do not map onto the brain were thrown out by filtering for voxels with notable activity, leaving primarily voxels that mapped onto person 1's brain. The result of this can be seen in figure 2. While it is possible to include multiple runs, or potentially even multiple people, this was not done primarily for time constraint reasons. Searchlight analyses are very computationally intensive (Etzel et al., 2013), and at the time of writing I did not have access to the computational resources necessary to complete a more complex analysis.

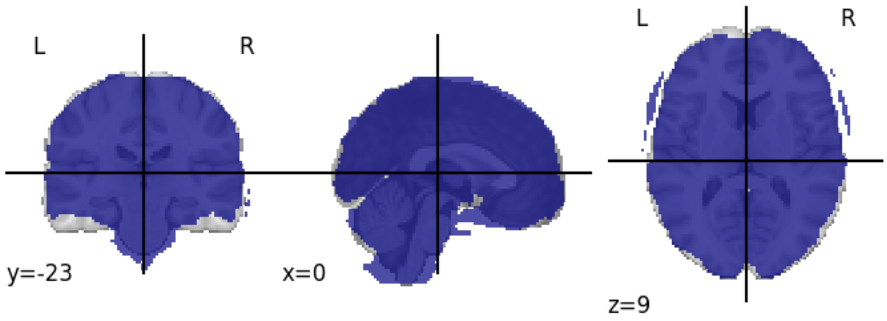

Figure 2: Figure depicting a mask of all active voxels in person 1's first run

Following this, a variable was created that contains information on which samples in the 4D Nifti contain congruent or incongruent data. To train the searchlight model, a cross-validation scheme must be used, in which I opted for Kfolds using the Python package sklearn's implementation, with 4 folds. This meant splitting the data into 4 folds, with each fold being used once as validation while (fold - 1) remaining folds form the training set. Following this, the analysis was performed, using the full brain mask, our 4D Nifti file, and the variable detailing the condition of our samples as the target for fitting the model.

4 Results

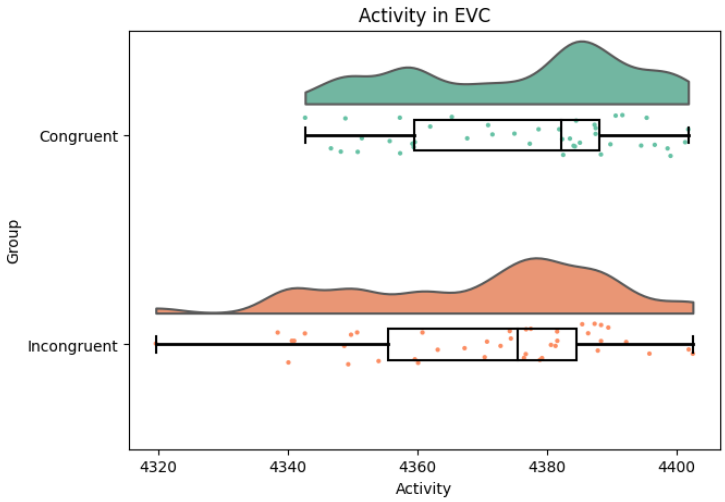

Figure 3: Raincloud plot of the ROI EVC, by congruent and incongruent

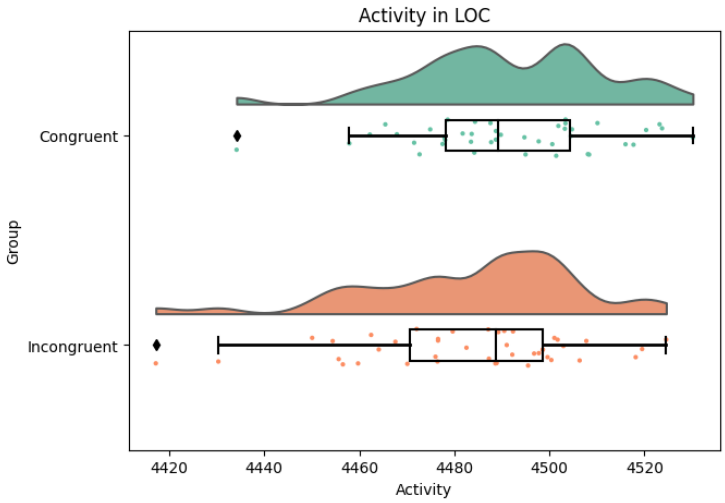

Figure 4: Raincloud plot of the ROI LOC, by congruent and incongruent

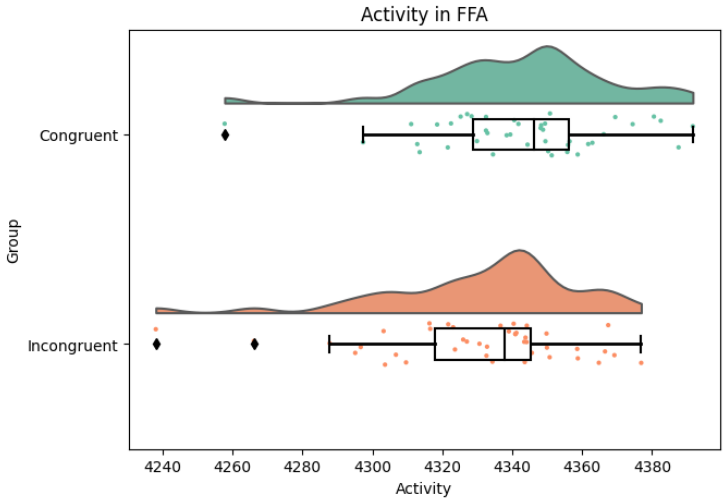

Figure 5: Raincloud plot of ROI FFA, by congruent and incongruent

In figures 3, 4 and 5 above, the raincloud plots are visible as a result of the univariate contrast analysis. Visible here is a noticeable difference between congruent activity and incongruent activity, with congruent activity being higher in all cases. This is counter to existing literature that predicts higher activity in situations where expectations are violated, meaning we should see a consistently higher activation in incongruent data rather than the congruent data (Puri et al., 2009) (Giesbrecht et al., 2006). The box plot reveals a systematic difference between congruent and incongruent activity across miniblocks, with the median and every quartile of the congruent activity means showing higher activity in every ROI. This is further explored in figure 6 below, confirming what we see here. The FFA ROI has some very strong outliers as well, which may be due to it being a relatively low number of voxels. Because of this, it is more subject to deviation. The FFA shows the strongest difference between congruent and incongruent as well, with its median being higher than the third quartile of the incongruent data.

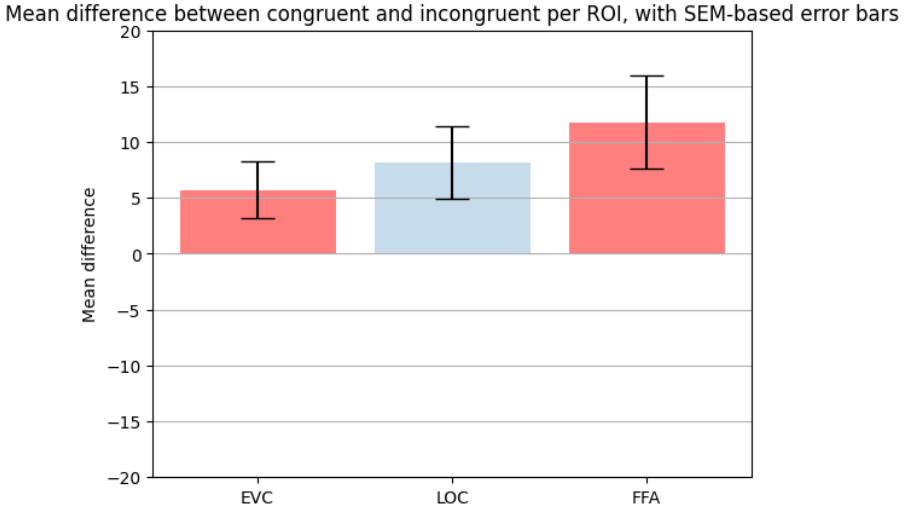

Figure 6: Bar plot depicting the differences between congruent and incongruent, by ROI, SEM error bars

ROI	EVC	LOC	FFA
Mean difference	5.70	8.16	11.78
SEM	2.54	3.25	4.17

Table 2: Table containing the mean differences and SEM by ROI

In figure 6 a plot depicting the findings of the differences analysis is shown. Table 2, a table containing the exact values that make up the plot is also shown. All error bars, which are based on the SEM (standard error of means), do not intersect with 0. This indicates that those areas respond quite strongly to stimuli, and are likely important for distinguishing congruency. The LOC and the FFA have higher mean differences, which likely means that there is a higher share of voxels within those areas that discriminate based on congruency.

Finally, all arrays of the differences had a one-sample t-test comparing to 0 performed on them to test for the statistical significance of the differences. This was done to determine whether the differences are systemic.

ROI	EVC	LOC	FFA
p-value	.0298	0.0151	0.0065

Table 3: Table containing the p-values of the one-sample t-tests performed on the differences, by ROI

As can be seen in table 3 above, all differences are significant compared to 0. This means that there is a systemic difference between miniblocks that contain congruent stimuli and miniblocks that contain incongruent stimuli, with congruent stimuli provoking a stronger response in all ROI than incongruent stimuli.

	Congruent values 1	Incongruent values 1	Congruent values 2	Incongruent values 2
Congruent values 1	1.000000	0.265495	0.819620	0.269937
Incongruent values 1	0.265495	1.000000	0.298059	0.855302
Congruent values 2	0.819620	0.298059	1.000000	0.296486
Incongruent values 2	0.269937	0.855302	0.296486	1.000000

Figure 7: Correlation matrix of 1 person, ROI EVC

In Figure 7 above, the result of the correlation analysis is shown. These correlate in the way we expect, with congruent data being more strongly correlated to other congruent data, and the same relationship shown for incongruent data. The difference between categories is calculated as follows: (con1 con2 + inc1 inc2) - (con1 inc2 + inc1 con2), with con1 referring to congruent values 1, inc1 referring to incongruent values 1, etc. The difference comes out to 1.107. To test for significance, repeating 10000 times, random splits of EVC congruent data and EVC incongruent data are taken, calculating score with the same formula as before. If the score is higher than that of the real data, a counter is increased. The counter comes out to 0 after running all repeats, providing statistical significance to the real difference as compared to random chance.

This proves that there is a statistically significant difference between congruent and incongruent activity from second to second, similarly to the univariate contrasts as seen in figure 6. However, from this correlation analysis we can conclude that there is a statistically significant difference in activation, even per second, during congruent stimuli and incongruent stimuli, which could be used to inform a learning model.

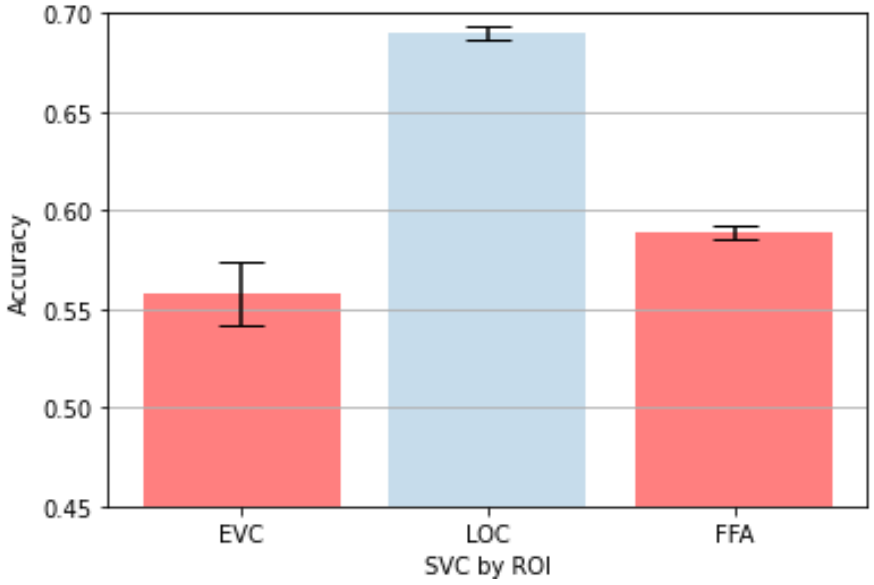

Figure 8: Accuracies of repeated SVC models, by ROI

Above, in figure 8, the accuracy values of the SVC algorithm are plotted. P-values for t-tests comparing accuracy values to random guessing (0.5 accuracy) are as follows: EVC SVC p-value = 0.0002254, LOC SVC p-value = 5.89e-67, FFA SVC p-value = 1.58e-40. All of these are statistically significant. It is unclear why the LOC performs better. Regardless, every SVC model can more accurately predict congruency than random guessing. These results indicate that there is information present in these regions of interest that can be used to determine congruency by learning models.

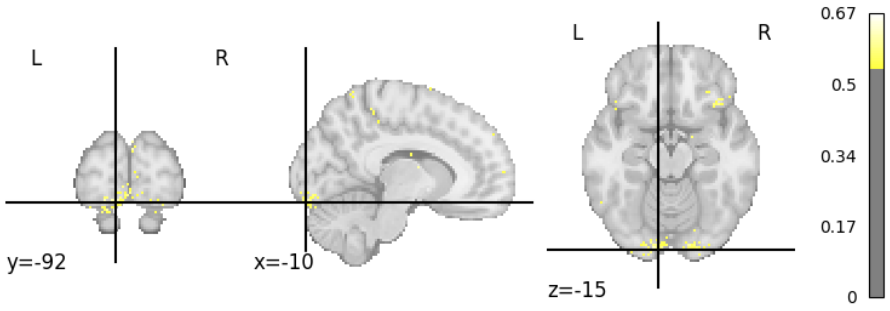

Figure 9: Plot of several slices of brain depicting voxels with >=0.55 accuracy

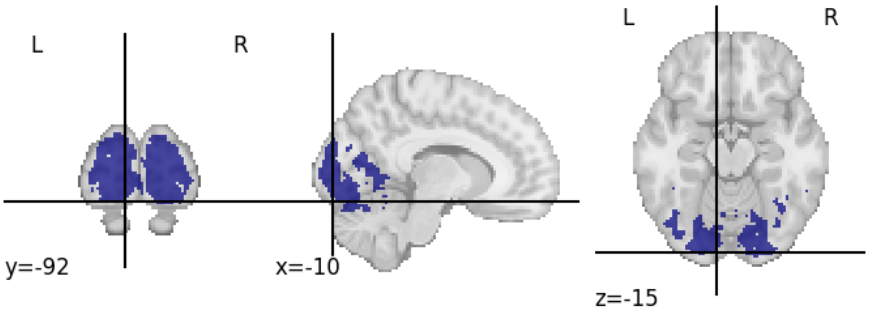

Figure 10: Plot of several slices of brain depicting voxels contained within the LOC, EVC and FFA brain regions for comparison with figure 9

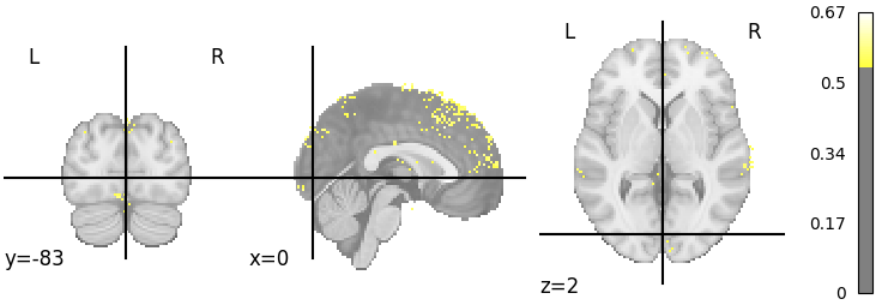

Figure 11: Plot of several slices of brain depicting voxels with >0.55 accuracy in different slices than 9

In the above three figures, figure 9, figure 10 and figure 11, a plot depicting some of the voxels that were the strongest discriminants for congruency are shown. Figure 10 serves as a point of comparison, as the blue areas depict the voxels contained in the EVC, LOC and FFA in the same slices of brain that the depiction in figure 9 has. Of note here is that most of the clusters that prove to be strong discriminants are roughly in the EVC and LOC regions, in the back of the brain. The clusters also appear to be bilateral, meaning there is less chance that these clusters are a result of noise. This means that the areas indicated by the clusters of voxels are active and a good discriminant for congruency. Some clusters in or near the prefrontal cortex are also visible, as well as near the center of the brain. The exact areas these clusters fall into are difficult to determine. However, because there are large clusters of voxels, and these clusters are again

largely bilateral, it is unlikely they are just a result of noise.

Within figure 11 more voxels can be seen in different slices of the brain compared to figure 9. Some are still present within the EVC and LOC; however, a majority of voxels are within the prefrontal cortex, and some are within the caudate nucleus as well. This indicates that these regions are also strong discriminators for congruency. It is likely that this is due to these regions being highly active in situations where expectation-violations occur, such as incongruency (Schiffer & Schubotz, 2011) (Danek et al., 2015).

5 Conclusion & Discussion

This paper set out to find the answer to two research questions. Primarily, is there a detectable, significant difference between brain activity while viewing congruent objects and while viewing incongruent objects - and if so, in which direction is this difference? Secondarily, using multivariate voxel analysis / searchlight analysis, can we find voxels that are strong discriminators for congruency, and where are they in the brain? Crucially, I believe I have found answers to these questions.

Firstly, by means of the correlation analyses, raincloud plots and SVC models, along with the relevant t-tests, we can be confident that congruent stimuli provoke a stronger response than incongruent stimuli. This is unexpected when we consider the literature. Notably, while Biederman et al. (1982) determines a detectability of differences in congruency, most literature that points to the EVC and LOC for object recognition and priming (Peelen and Kastner (2011) and Puri et al. (2009)) suggest that expectation-violations (incongruency) heightens neural response. In this thesis I have found the opposite, with congruent stimuli systematically generating a stronger response across all regions of interest. The LOC especially aids in size perception (Cate et al., 2011), which is expected to be important for congruency, as congruency in this context is an expectation-violation of size (Wang et al., 2004). What is unexpected however is that the LOC still strongly works as discriminator when the size of the object on the retina is not different, but the object size appears smaller or larger than expected, as in incongruent compared to congruent. The importance of the LOC in the task outlined in this thesis is strongly reflected in the more sensitive analysis provided by the SVC, with a model being trained off of the LOC obtaining the highest accuracy score of all regions of interest. What is somewhat unexpected based on the literature is how strongly FFA works as a discriminator of congruency. While literature suggests it is a predictor of object recognition (McGugin et al., 2016), it was a better discriminator than the EVC in all analyses, providing the strongest differences in univariate contrast analysis and visually within raincloud plots, and performing better than the EVC when used to train an SVC model. This suggests that the FFA is of high importance for determining congruency, based on the results. This could be the result of the proximity of the fusiform face area to the lateral occipital cortex, with voxels within the FFA being influenced by the LOC.

Secondly, the findings based on the searchlight analysis are largely in line with established literature. Some of the strongest clusters of voxels that serve as discriminants for congruency are found in and around what appears to be the LOC and EVC areas. As these two areas aid in extracting local features of shape and size, and are both very involved in object recognition, it was expected that these areas would serve as discriminators. (Grill-Spector et al., 2001) (Lee, 2003) (Gayet & Peelen, 2022)(Khateb et al., 2002). As outlined in the previ-

ous paragraph however, the LOC was of surprising importance. Moreover, a new finding is the strength of congruence discrimination shown by voxels in the prefrontal cortex and caudate nucleus areas. Previous literature established the importance of these regions during expectation-violations (Danek et al., 2015) (Schiffer & Schubotz, 2011) and perceiving object size (in the case of the prefrontal cortex) (Perini et al., 2020). However, the importance of these regions specifically in determining congruency as a violation of size-context relationships (Biederman et al., 1982) has not been established prior to this analysis and would be worth examining in further research. These areas are strongly represented in the searchlight analysis, suggesting that processing expectation-violations like incongruency is an important step for visual processing even in areas less commonly associated with object recognition. Concluding, while some areas that are highly active during object recognition do play a large part during the tests with varying congruency, the prefrontal cortex, caudate nucleus, and potentially other areas are also highly active and discriminatory on congruency. This suggests involvement of a large part of the brain during object recognition and expectation-violation, and not the compartmentalised view of this process as older literature suggests (Marr, 1982).

There are several limitations in this study, largely stemming from a lack of quantity of data and lack of an expert in relative domains to explain quirks in the data. First off, I believe that the SVC model can be trained to achieve higher accuracy than what was seen within the results of this thesis, which may be a goal in future research. While in its current form the SVC models are sufficient to answer the research questions, they are not highly accurate. To increase accuracy, more runs may be performed, more extensive models of cross-validation and hyperparameter optimisation may be considered, among other options. These can be considered in future research where higher accuracy is a goal.

Lastly, the searchlight analysis was limited in its scope and data usage. Quite simply, because I do not have the computational resources or the time required, only a single run of data was considered. Also due to time constraints no formal test for significance was performed that would account for the false discovery rate that is present in searchlight analyses (Etzel et al., 2013). Each time the analysis was run, it involved approximately a day of running time and significant computational resources, even in its current limited form. Further, no cluster permutations were performed to filter out noise, nor corrections for multiple comparisons, also because of the limited time-frame of this analysis. If further research is to be done, significantly time and resources would be required. Ideally, same as with the other analyses, more runs and participants could then also be considered. That said, I believe it is unlikely that the findings present in the searchlight analysis are due to noise, as the found voxels cluster closely and are bilaterally present. In this further research, a neuroimaging expert would be well-served for determining in which areas all the discriminatory voxels lie, and what this could mean for the process of object recognition and processing

size-distance violations.

In conclusion, the strongest findings in this thesis are that congruent stimuli produce a stronger response in the EVC, LOC and FFA areas of the brain compared to stimuli containing incongruently placed and sized objects. Furthermore, based on searchlight analysis, I found that visual expectation-violations can be discriminated as occurring or not occurring in areas commonly associated with visual processing such as the LOC and the EVC, as well as in areas that have previously been shown to activate strongly during expectation-violations such as the caudate nucleus and prefrontal cortex. Lastly, based on the strong discriminatory ability of the FFA, the FFA plays a large role in object recognition during tasks involving visual expectation-violations, more than previously recognized.

References

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology*, *14*(2), 143–177.

Cate, A. D., Goodale, M. A., & Köhler, S. (2011). The role of apparent size in building-and object-specific regions of ventral visual cortex. *Brain research*, *1388*, 109–122.

Danek, A. H., Öllinger, M., Fraps, T., Grothe, B., & Flanagin, V. L. (2015). An fmri investigation of expectation violation in magic tricks. *Frontiers in psychology*, *6*, 84.

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in cognitive sciences*, *22*(9), 764–779.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, *100*(3), 363.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual review of psychology*, *47*(1), 273–305.

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: Promise, pitfalls, and potential. *Neuroimage*, *78*, 261–269.

Friman, O., Borga, M., Lundberg, P., & Knutsson, H. (2002). Exploratory fmri analysis by autocorrelation maximization. *NeuroImage*, *16*(2), 454–464.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123–144.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects. *Nature neuroscience*, *2*(6), 568–573.

Gayet, S., & Peelen, M. V. (2022). Preparatory attention incorporates contextual expectations. *Current Biology*, *32*(3), 687–692.

Giesbrecht, B., Weissman, D. H., Woldorff, M. G., & Mangun, G. R. (2006). Pre-target activity in visual cortex predicts behavioral performance on spatial and feature attention tasks. *Brain research*, *1080*(1), 63–72.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research*, *41*(10-11), 1409–1422.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and brain sciences*, *27*(3), 377–396.

Khateb, A., Pegna, A. J., Michel, C. M., Landis, T., & Annoni, J.-M. (2002). Dynamics of brain activation during an explicit word and image recognition task: An electrophysiological study. *Brain topography*, *14*(3), 197–213.

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology-Paris*, *97*(2-3), 121–139.

Marr, D. (1982). Vision wh freeman and company. *San Francisco*, 41–98.

McGugin, R. W., Van Gulick, A. E., & Gauthier, I. (2016). Cortical thickness in fusiform face area predicts face and object recognition performance. *Journal of cognitive neuroscience*, *28*(2), 282–294.

Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences*, *108*(29), 12125–12130.

Perini, F., Powell, T., Watt, S. J., & Downing, P. E. (2020). Neural representations of haptic object size in the human brain revealed by multivoxel fmri patterns. *Journal of neurophysiology*, *124*(1), 218–231.

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain research*, *1301*, 89–99.

Rolls, E. T. (2001). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Vision: The Approach of Biophysics and Neurosciences*, 366–395.

Schiffer, A.-M., & Schubotz, R. I. (2011). Caudate nucleus signals for breaches of expectation in a movement observation paradigm. *Frontiers in Human Neuroscience*, *5*, 38.

Wang, S.-h., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition*, *93*(3), 167–198.

6 Appendix

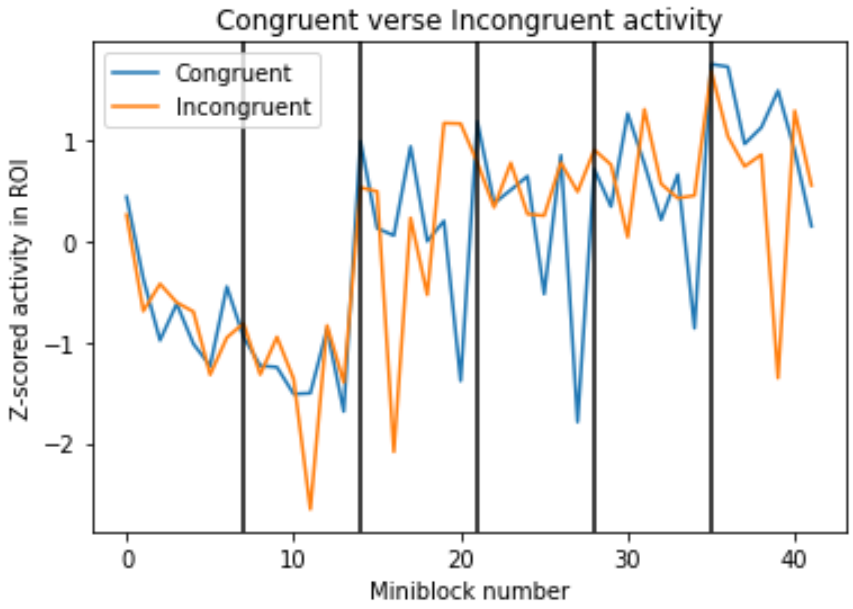

Figure 12: Z-scored activity in ROI EVC & LOC, with congruent plotted against incongruent, single person and all 6 runs

While this graph did not suit the paper in its current form, and the visualisation of it is not ideal, it is included here to show the variation of activity even within a single person. Every vertical line is a new run, and as such this graph is not a proper time series line graph. It does show a potential future challenge, however. The dips are not due to the blocks in which the participant sees no stimulus, as these were taken out, and the cause of these is unknown. This is an example of the utility of a domain expert in analyzing this data and explaining quirks in the data collection.

	Congruent values 1	Incongruent values 1	Congruent values 2	Incongruent values 2	Congruent values 3	Incongruent values 3	Congruent values 4	Incongruent values 4
Congruent values 1	1.000000	0.991189	0.590176	0.590182	0.522384	0.521149	0.430240	0.431890
Incongruent values 1	0.991189	1.000000	0.591053	0.590686	0.522838	0.521611	0.428709	0.430223
Congruent values 2	0.590176	0.591053	1.000000	0.994453	0.491361	0.492041	0.355828	0.355900
Incongruent values 2	0.590182	0.590686	0.994453	1.000000	0.491931	0.492481	0.355493	0.355978
Congruent values 3	0.522384	0.522838	0.491361	0.491931	1.000000	0.992052	0.317953	0.318530
Incongruent values 3	0.521149	0.521611	0.492041	0.492481	0.992052	1.000000	0.316481	0.317251
Congruent values 4	0.430240	0.428709	0.355828	0.355493	0.317953	0.316481	1.000000	0.981756
Incongruent values 4	0.431890	0.430223	0.355900	0.355978	0.318530	0.317251	0.981756	1.000000

Figure 13: Correlation analysis between participants, split by congruency, EVC area

This is the result of an older correlation analysis I performed, testing correlations between participants. As the conclusions that can be drawn based on this matrix are obvious to most within the field, being that between-subject analysis is infeasible, it was excluded from the main text. However, it was informative for me as I was learning about the data while I did not have a clue of it yet, so it is included here.