Project Epsilon progress report
The Neural Basis of Loss Aversion in Decision-Making Under Risk

Jo, Min Gu	Kaam, Soazig	Li, Zhuangdi	Yu, Timothy	Zhi, Ye
mingujo	soazig	lizhua	timothy1191xa	ye-zhi

December 14, 2015

1 Abstract

Please update this paragraph

2 Introduction

The study *Neural Basis of Loss Aversion in Decision-Making Under Risk* [**?**] focuses on decision-making process, especially on the correlation between the neural activity and the reluctance to lose. 16 people were presented 255 gambling situations with a 50% of success. Each situation was associated with a potential gain and loss that were randomly selected. The gains were ranging from $10 to $40 while the losses from $5 to $20. The participants were asked to assess their level of willingness to accept or reject the gamble using a 4-point likert scale [1: strongly accept, 2: weakly accept, 3: weakly reject, 4:strongly reject]. The response time was also recorded for each case. The imaging data were collected using the fMRI method. They were processed and analyzed in order to identify the regions of the brain activated by the decision making process. This study also investigated the relationship between the brain activity and the behavior of the subjects towards the gambling situations using a whole-brain robust regression analysis.

3 The data

The data we are using can be found on the OpenfMRI website at the following address: `https://www.openfmri.org/dataset/ds000005`, the dsnum is ds005. For our project, we are specifically using the behavior data and the BOLD data that are organized.

For each of runs per subject (3), the behavior data contains the timestamp of each survey question (onset), the gain/loss combinations (gain and loss), the response for the particular trial (respnum) from the 4-point likert scale. The researcher created a response category (respcat) to be used in their binary choice model that combines the "reject" answers together on one hand and the "accept" answers together on the other hand. BOLD data contains compressed 4-dimensional brain images for each subject's run. The folder also comport Quality Assurance (QA) files and a report.

4 Behavioral Data Analysis

4.1 Introduction

First of all, we generated some summary statistics including correlation among variables, and tried both linear and logistic regression analysis for behavioral data. The scientific questions that we have are - if we can predict response time and response (to gamble or not to gamble) based on the gain and loss.

4.2 Methods

We did some explanatory data analysis and regression analysis using behavior data. For explanatory data analysis, we generated some summary statistics, including correlation among variables and simple plots to better understand the behavior. And then we used regression analysis to mainly answer two scientific questions. The scientific questions that we have are:

- If gain/loss would be significant for individuals who choose to participate and how much time it would take for them to respond.
- If gain/loss would be significant for whether individuals would like to participate in the gamble.

4.2.1 Linear Regression

I will change this.

1. Response Time $\sim$ gain + loss
2. Response Time $\sim$ diff(gain-loss)
3. Response Time $\sim$ ratio(gain/loss)

4.2.2 Logistic Regression

To answer the second question, we fit logistic regression between Accept/Reject Gamble and gain/loss. According to our analysis, the decision to whether take the gamble of most of subjects, in general, is more affected by loss amount rather than by gain amount. For example, below is the analysis on the subject 3. The regression line shows that it well follows the border between the two decisions: 1 (gamble) and 0 (not gamble). Right side of the line illustrates the decision to not gamble and it takes up more area relative to the opposite decision.

L ogistic Regression is a statistical technique capable of predicting a binary outcome. Since, in this data, the researchers classify the decision to gamble as '1' or '0' otherwise, we can use logistic regression technique to explain the subject's tendency to gamble or not based on the condition of the gain and loss amount given in the process of experiment. Our goal is to identify the how gain and loss amount influence each subject's response. To do this, we use the statsmodels Logit function. We specify the response column in the behavior txt file as the one containing the variable we're trying to explain and the gain and loss columns as the predictor variables. After plotting the results on the plot, we were able to see some interesting behaviors of some subjects. As you see from the plot , Subject 1 is in general more risk seeking: as long as the gain amount is large enough as 20 dollars, he decides to gamble. However, Subject 3 shows the opposite behavior: she does not participate in the gamble when her loss amount is higher than 10 dollars no matter what the gain amount is. (To see the overall behaviors from all subjects, see the appendix) Overall, we could see that the logistic regression line fits well on the border between the decision to gamble and not gamble.

4.3 Results

For linear regression, ratio is a significant predictor and people would actually care more about loss than gain.

T he paper illustrates as ?people typically reject gambles that offer a 50/50 chance of gaining or losing money, unless the amount that could be gained is at least twice the amount that could be lost (Sabrina)?. In the experiment, the given gain and loss amount ratio to each subject is around 2 to 1. This refers subjects would not merely show risk averse behavior every trial. We could confirm this trend by observing the plots. We are hard to tell whether subjects are risk averse or not.

5 Image Data Analysis

5.1 Introduction

After exploring behavior data, we proceed to image data analysis. First we need to apply convolution to connect behavior stimuli and neural activity. Then we can run general linear regression to find activated voxels across time course. Using hypothesis testing, we can actually locate and visualize the activated voxels. After finishing basic steps, we try to apply noise modeling and PCA to compare the MRSS so that we can finally decide our design matrix.

5.2 Methods

Here are some of the methods (this needs to be updated soon as well.)

5.2.1 Convolution

Our experiment is event-oriented. The subject is shown with the different conditions such as gain and loss amounts over random time. After being provided with the conditions, the blood flow responses starts to fluctuate. To predict the fMRI signal to an event, we need to predict the hemodynamic responses to neural activity. A predictor neural time course is typically obtained by convolution of a condition of the experiment with a standard hemodynamic response function. With this predictor, we build our design matrix for our general linear model on each voxel of a subject's brain. To produce such predictor, we practiced two different approaches.

- Convolving with canonical HRF
  A typical BOLD response to a single, impulsive stimulation resembles a linear combination of two Gamma function. This would model a signal is instantly at its peak level at the onset of a given condition and instantly undershoot back to around baseline with the offset of a condition. We can use this hemodynamic response function as a canonical one. Generally, the canonical HRF should be a good fit if we believe the subjects to be normal in many cortical regions. Using this canonical HRF will help us to find how much the canonical HRF has to be scaled enough to account for the signal. However, we want to be more in detail as long as the onsets of the HRF can happen in the middle of volumes due to the conditions given at different times. The amplitudes vary according to the parametric gain and loss conditions. Thus, the true shape of HRF for each subject should vary.

- Convolving at finer time resolution
  Therefore, we would make a neural and hemodynamic regressor at a finer time resolution than the TRs, and later sample this regressor at the TR onset times. This refers that stimulus onsets do not have to be synchronized with scan TRs.

5.2.2 GLM

The first matrix we get from convolution has five columns, which correspond to a column of ones and 4 cond.txt files in our dataset, respectively. After we get the convolution matrix, we use it as our design matrix to run the generalized linear regression on the image data. The dimension of our data is (64, 64, 34, 240), so, first we reshape our data into 2 dimensional array, which has the shape of (64*64*34, 240); the first dimension corresponds to 3-dimensional voxel indices and the second dimension corresponds to the time slice. Then we pass our design matrix into the glm function to calculate the related beta hats. Thus, there are in total 139624 beta hats that we get from the regression correspond to the first three dimensions of our image data. For example, the first beta hat contains the information about the voxel (0,0,0). Then we turn the beta hats back into 4-dimensional shape and run the diagnostic functions on the 4-d beta hats. Based on the predictors, we can calculate the fitted values and then the residuals. We use

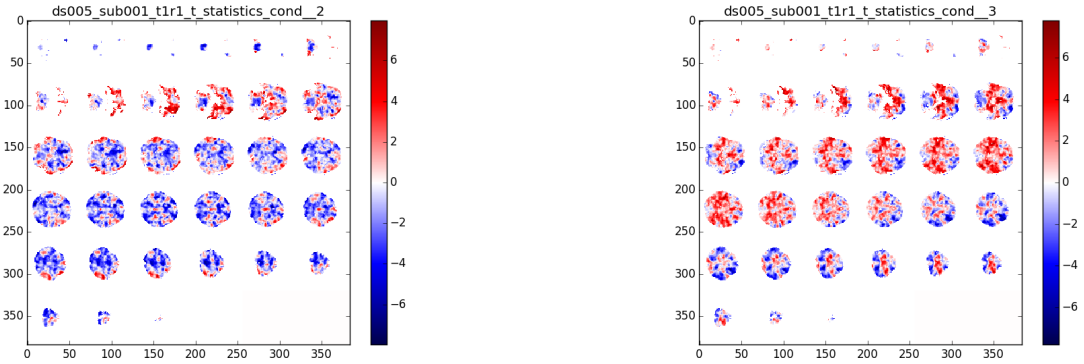
the MRSS of the first three dimensions as a measurement of our regression; in general, a smaller MRSS indicates a better performance of the regression model.

5.2.3 Smoothing

After we tried with the normal convolution matrix, we also generated high resolution convolution matrix and used it for linear regression. It turned out that the MRSS is just reduced by a little bit. Then we write a smoothing function to implement the multidimensional Gaussian filter on our data. We repeat the same procedures as what we have done in normal convolution on the smoothed data and the MRSS are reduced sharply. Therefore, we concluded that the smoothing method is a good pre-processing when we do the linear regression.

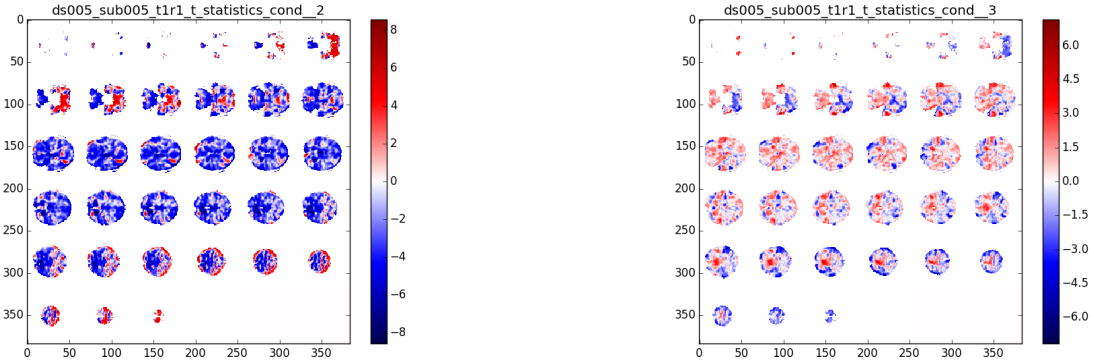
5.2.4 Hypothesis Testing

From linear regression, we can get t-statistics for different conditions(task on/off, gain, loss, distance).For each condition, we will have a 3D t-statistics matrix. For visualization, we first added mask based the mean voxel and the histogram. We set a boolean mask which takes larger than 375. Also we used smooth function and better color txt to generate a better image. Then we plotted the t statistics map for gain/loss.



(a) Gain (b) Loss

Figure 1: t statistics map for gain/loss (subject1)



(a) Gain (b) Loss

Figure 2: t statistics map for gain/loss (subject5)

The larger the t-statistics, the more significant. Thus the red spots represents the activated voxels for gain and loss. For subject 1, gain has more activated voxels. However, for subject 5, loss has more activated voxels.

5.2.5 Noise Model

5.2.6 PCA

5.3 Results

In terms of convolution, to analyze the difference between two approaches, we compare the MRSS from two linear regressions on image data of three subjects (1,2,3) using convolution predictors from two different approaches. In the below table, we see the MRSS from linear regression using the latter approach has slightly lower residuals compared to the former method. This makes sense because, using the latter method, we are able to more elaborately preprocess the data.

6 Multi-Comparison

6.1 Introduction

As our intended goal of this project is to locate the ROI $RegionofInterest$ in a brain according to this mixed gamble task, the significant issue is the specification of an appropriate threshold for statistical maps. Therefore, we are interested in multi-comparison across subjects. In the previous analysis on single subject, we were able to locate the generally activated voxels. With these data, we attempted to find out the general pattern in each voxel of brains across 16 subjects.

6.2 Methods

As our intended goal of this project is to locate the ROI $RegionofInterest$ in a brain according to this mixed gamble task, the significant issue is the specification of an appropriate threshold for statistical maps. Therefore, we are interested in multi-comparison across subjects. In the previous analysis on single subject, we were able to locate the generally activated voxels. With these data, we attempted to find out the general pattern in voxels of brains across 16 subjects.

To multi-compare, we chose to explore on filtered data set (shape: 91*109 * 91) since original data set is not normalized in terms of voxel location. Applying our general linear modeling on each subject as above, we first collected all beta values for each voxel across time course of each subject. Now, we have beta values of shape of (91, 109, 91) for each of 16 subjects. We compare on each single voxel across subjects. To do this, We have total 4 steps.

- Calculate average of beta value on a single voxel across 16 subjects. Do this for whole voxels.

- Calculate standard deviation of beta value on a single voxel across 16 subjects. Do this for whole voxels.

- Calculate T-statistics of beta value on a single voxel across 16 subjects. Do this for whole voxels.

$$t - statistics = \frac{mean}{\frac{SE(mean)}{\sqrt{n}}}$$

- Calculate P-value of beta value on a single voxel across 16 subjects. Do this for whole voxels.

For the 4th step, we have an issue of false positives: since we're executing 9110991 number of hypothesis testing, note that an error probability of p value = 0.05 means that if we would repeat the same test 1000 times and assume that there is no effect of the experiment, we would wrongly reject or accept the null or alternative hypothesis on average. Therefore, we decided to strengthen our p-value threshold with Bon-ferroni correction: our new p-value threshold is 0.05/9110991

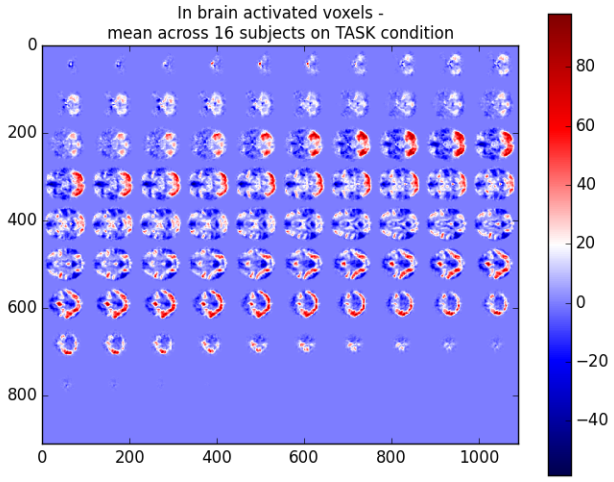


Figure 3: Mean beta values on each voxel across 16 subjects on TASK condition

6.3 Results

As the image of mean of beta values on each voxel across all subjects on task condition indicates, we are able to classify which portion of brain is in general activated due to this experiment. However, there should be fluctuation over different subjects. To identify this, standard deviation is also shown. As the image in the appendix illustrates, fluctuation occurs on similar regions because each subject's hemodynamic response degree must be different. We found out that the t-statistics in general are not significant enough; therefore, we could not elaborately indicate the specific portion of brain related to the mixed gamble task. However, t-statistics and p-value of voxels over a brain well correspond to each other.

7 Discussion and Conclusion

7.1 Discussion

7.2 Conclusion

Update the conclusion!

8 Appendix

8.1 Resource1

8.2 Resource2

8.3 Resource3