The Neural Basis of Loss Aversion in Decision-Making Under Risk

Chang, Siyao	Gong, Boying	Hsieh, Benjamin	Qiu, Brian	Zhu, Jiang
`changsiyao`	`boyinggong`	`BenjaminHsieh`	`brianqiu`	`pigriver123`

November 30, 2015

Abstract

Our paper is the *Neural Basis of Loss Aversion in Decision-Making Under Risk* [**?**]. The experiment investigates the phenomenon of loss aversion - where individuals decisions are influenced by the amount of potential loss more than they are by the amount of potential gains.

Our analysis of the paper aims to verify this result and to replicate the connections found between neural loss aversion (shown in brain activation) and behavioral loss aversion (shown in the actual decisions to accept or reject gambles based on potential gain/loss). Our work has found that the neural loss aversion is [placeholder] and that the behavioral loss aversion is [placeholder]. Our work also shows that neural and behavioral loss aversion are [placeholder], and that female and male subjects exhibit [placeholder].

1 Introduction

The experiment conducted in the paper itself involved giving 16 subjects a total of 256 combinations of gain/loss in dollars with a 50 percent chance of winning. The subject's decision of whether to accept or reject each proposed gamble was recorded as well as the their brain activity in the fMRI machine as they made their decision. The neural activity is measured in BOLD signals.

We determine the beta coefficients for each voxel in the subject's brain based on a linear regression on the BOLD signal data. We use these coefficients as evidence for increasing activation in certain voxels in the brain when the subject is making a decision to accept or reject, and so in this way we have a proxy for which regions of the brain are involved with loss-aversion decision-making. We also determine the p-values to validate these results.

We also run a logistic regression to determine each subject's willingness to accept or reject based on the values of potential loss and gain. This is the measure of behavioral loss aversion, and in the end we correlate this with ou neural loss aversion, making it clear that increasing gains correspond to increasing activity in certain parts of the brain.

We also aim to investigate briefly if there are any statistically significant variations between the male and female subjects in addition to showing the brain activations through various graphs.

2 Data

2.1 Overview

The study used 16 right-handed, healthy, English-speaking participants recruited through ads posted on UCLA. Out of 16 subjects, 9 were female and the mean age was 22 $\pm$ 2.9 years. [**?**]

2.2 Behavorial Data

The behavioral data consists of each subject undergoing 3 trial runs for the "gamble" task, in which each subject is presented with a combination of potential monetary gains and losses given a 50/50 chance of to win/loss. Each trial run consists of 86 different combinations of rewards/penalities spread out accross 474 seconds. Intervals between each onset of task range from 4 to 8 seconds. Subjects were given the 4 choices in reponse to each gambling proposal:

1. Strong Accept
2. Weak Accept
3. Weak Reject
4. Strong Reject

The choices are recorded by denoting reponse numbers 1, 2, 3, and 4, respectively. Furthermore, the response time for each gambling decision was recorded in seconds.

2.3 BOLD Data

Blood-oxygen-level dependent (BOLD) imaging data were collected from each subject as he/she performed the gamble tasks. 240 time scans were done on each run with a time between each scan of 2 seconds. So total scanning time is 480 seconds. Each scan consists of a snapshot consisting of a64 by 64 by 34 image matrix.

There are also 4 model conditions, with events corresponding to

1. Task
2. Parametric Gain
3. Parametric Loss
4. Distance from Indifference

2.4 Processing

Before we run our analyses and fit betas to our data, we first take a few steps to clean the BOLD datasets. We graph the dvars (RMS of the signal derivatives) and the framewise displacement, and we use that in conjunction with the mean signal of the BOLD data to determine outliers to remove. The BOLD data is smoothed spatially in order to make clearer the signal in relation to the noise present. We also take the mean signal of the BOLD data across the run and plot the histogram for each run and subject to help manually determine a good threshold for a mask we use to isolate more active voxels in the brain. This helps us find beta coefficients for more relevant voxels. We also model and remove the linear and quadratic drift that may be present in the runs. We use subject 2 run 2 to show an example of our outlier results.

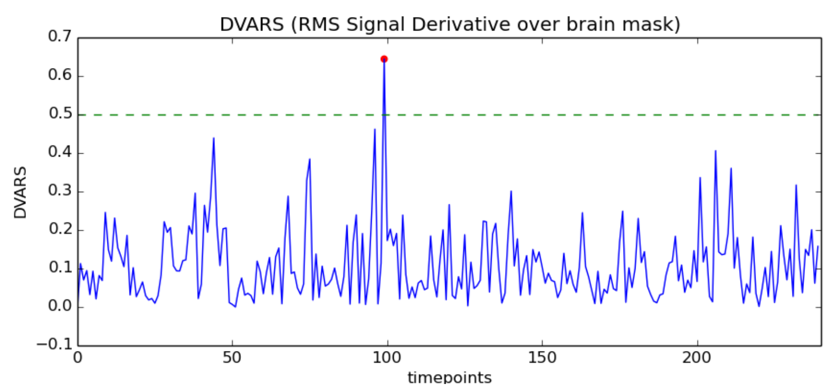


Figure 1: DVARS for subject 2 run 2

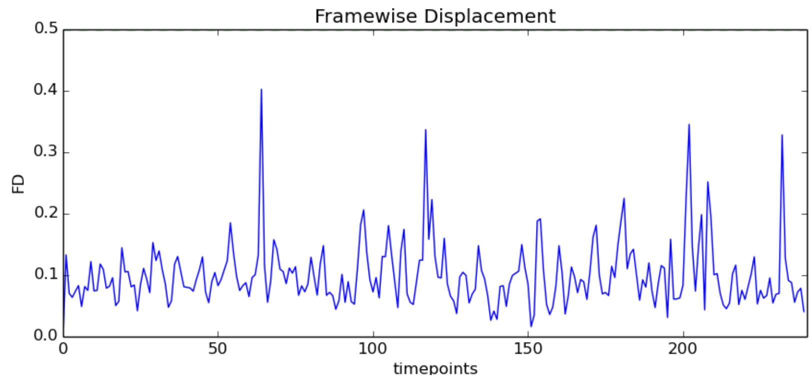


Figure 2: Framewise Displacement for subject 2 run 2

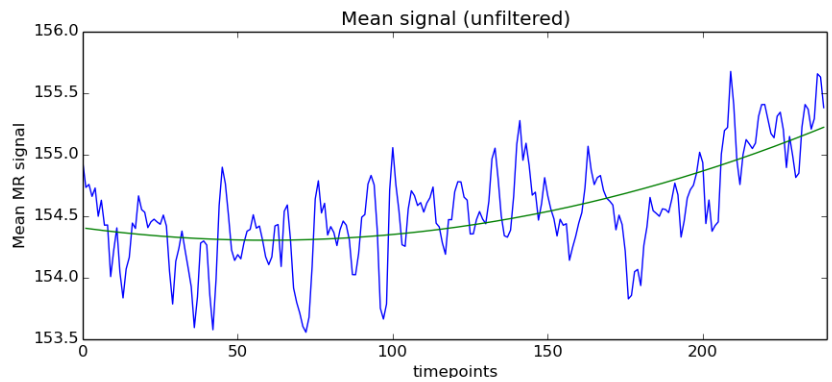


Figure 3: Data mean for subject 2 run 2

3 Methods

3.1 Models and analysis

In this section, we present the models we used to find the relationship between behavioral and neural loss aversion cross participants as well as how participants react to different loss and gain levels. For behavioral data, we fit the logistic regression models for each subject and use the coefficients of loss and gain to calculate the behavioral loss aversion levels. For neural data, we fit both linear multiple regression models and mixed-effects models in order to collapse three runs for each subject into one model. We analysis the fMRI data use both models and compare the results we obtained. Neural loss aversion levels are calculated using the coefficients of loss and gain.

3.1.1 Behavioral analysis

We fit a Logistic regression model on the behavioral data to examine how the response of individuals relates to the size of potential gain and loss of a gamble. Originally there are four acceptability judgments categories, here we collapse the categories into binary response(accept/reject). Following is the model:

$$\log \frac{p(X)}{1-p(X)} = \beta_0 + \beta_{loss} * X_{loss} + \beta_{gain} * X_{gain} \tag{1}$$

where X_{loss} and X_{gain} are the potential loss and gain value separately, Y_{resp} is a categorical independent variable representing the subjects' decision on whether to accept or reject the gambles:

$$Y_{resp} = \begin{cases} 1 & \text{If the subject accepted the gamble.} \\ 0 & \text{If the subject rejected the gamble.} \end{cases}$$

Then we calculate the the behavioral loss aversion (λ) for each subject as follows, note that for simplicity, we collapse 3 runs into one model for each participant.

$$\lambda = -\beta_{loss}/\beta_{gain} \tag{2}$$

We use λ as the metric for the degree of behavioral loss aversion for each participant.

3.1.2 Linear Regression on fMRI data

For each voxel i, we fit a multiple linear model:

$$Y_i = \beta_{i,0} + \beta_{i,1} * X_{ldrift} + \beta_{i,2} * X_{qdrift} + \beta_{i,loss} * X_{loss} + \beta_{i,gain} * X_{gain} + \epsilon_i \tag{3}$$

where Y_i is the BOLD data of voxel i, X_{ldrift} and X_{qdrift} are linear and quadratic drift terms. For each voxel, we calculate the neural loss aversion η_i:

$$\eta_i = (-\beta_{loss}) - \beta_{gain} \tag{4}$$

Using the voxelwise neural loss aversion, we do a region-specific analysis on BOLD data for each participant. That is, we plot a heat map of η_i and $\beta_{i,loss}$, $\beta_{i,gain}$ for each participant to find out the regions with significant activation and regions which show a significant positive or negative correlation with increasing loss or gain levels.

3.1.3 Mixed-effects model on fMRI data

The fact that we have 3 runs of data for each participants leads us to consider using mixed effects model to analysis the data set. For each voxel i, we fit the following mixed-effects models, note that here we only include the intercept term for random effects.

$$Y_{i,k} = \beta_{i,0} + \beta_{i,1} * X_{ldrift} + \beta_{i,2} * X_{qdrift} + \beta_{i,loss} * X_{loss} + \beta_{i,gain} * X_{gain} + \gamma_{i,k} + \epsilon_{i,k}, \quad k = 1, 2, 3 \tag{5}$$

Then we calculate the neural loss aversion level and plot heat maps same as the about section for multiple linear regressions and compare the results of two models.

3.1.4 Whole brain analysis of correlation between neural activity and behavioral response across participants

We then apply the above model on the standard brain to analysis the neural activity and behavioral response across participants. For each participant, we pick up several regions with highest activation level, calculate the mean neural loss aversion $\bar{\eta}$ within these specific region. Thus we could examine the relationship between neural activity and behavioral using the following regression model:

$$\lambda = \alpha_0 + \alpha_1 * \eta + \epsilon \tag{6}$$

where the sample size is the number of participants(16).

3.1.5 Cross-validation

We fit linear models for each voxel for each participant. For each linear model, we do a k-fold cross-validation. Since the sample size for each linear regression model range from 80-90, we choose to use 10 fold cross-validation, which means the original sample is randomly pertitioned into 10 equal sized subsamples.

In the behavioral analysis using Logistic regression, since the responce variables are binary, we calculate the misclassification error rate to summarize the fit. In the neural linear regression model using BOLD data, we use the mean squared error to summarize the errors.

3.1.6 Inferences on Data

After fitting regression models on our BOLD and behavioral data, we would try assessing and validating our models. In order to do this, we would calculate for the residual sum of squares for our model. We have to do three tests for the model. The first one is that we calculate the t-statistics and p-value for our beta coefficients to check whether our beta parameters are statistically significant at a significance level of 5%. The second one is that we calculate the residuals of this linear model and check whether it follows a normal distribution. The third one is that we calculate the R-Squared value and the adjusted R-squared value to see whether the values are good for the linear regression model.

3.2 Explanation on model simplification

3.2.1 Use of Data

First of all, for simplicity reasons, we are not using all the regressors the paper used. The model in the paper performed regression on the BOLD data with gain, loss and euclidean distance to indifference. In our model, we are leaving out the regressor euclidean distance to indifference. The paper and its supplement material didn't document the exact way the authors calculated this parameter; we are having a hard time reproducing this parameter. Therefore, we decide to leave out this parameter when doing our own regression.

3.2.2 Perform multiple regression and mixed-effects model on BOLD data

We fit two separate models on neural data. In the original data analysis, the authors performed a mixed effect model when regressing the potential gain and loss values against the BOLD data across runs, since there are three different runs for each subject and the authors were trying to incorporate all three runs into one model. The mixed effect model adds a random effects term, which is associated with individual experimental units drawn at random from a population. In this case, it measures the difference between the average brain activation in run i and the average brain activation in all three runs.

While fitting the mixed-effects model using package *statmodels*, we are simplifying the model because it is much easier to perform a simple linear regression in python. For the multiple regression part, we write functions to calculate coefficients and do multiple inferences on the regression model.

After looking at the initial result from our linear regression model, we can decide whether we want to further explore the relationship between the dependent variable (BOLD data) and the independent variables (gain and loss).

3.3 Model Diagnostics

By using a linear regression model for calculating the loss and gain β values in the neural (BOLD) signals, we have made the following underlying assumptions:

1. Normailty of errors
2. Homoscedasticity (constant variance) or errors
3. Linearity of relationship between the explanatory and response
4. Statistical independence of errors
5. All voxels and all subjects follow the same linear model.

The first four assumption can be motivated by using the residuals as a proxy for the errors since the errors are unobservable. Here we check the first three assumptions by following the standard regression diagnostics plots and analysis. The fourth condition, checking for independence of errors, is equivalent to checking that no correlation exists between consecutive errors in our time serires structure. A simple explanation is given in *Simplifications of Model*, so it is leaved out here. For the fifth assumption, it may that there are some observations/conditions that do not obey the linear model. The methods used to detect these usual observations is further discussed in *Outlier Detection*, so it will be left out here.

3.3.1 Linearity and Constant Variance

The residual versus fitted values plot ?? is a standard plot for observing constant variance of errors and potential deviations from linearity

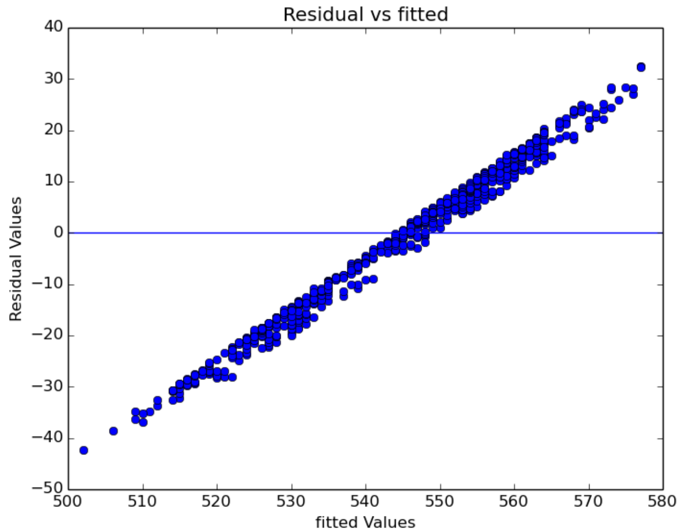


Figure 4: Residual vs Fitted Values for voxel 40000 across time

From the figure we can see there is clear linear pattern to residual. The residuals seems to increase as the fitted values increase, suggesting a variance-stabilizing transformation to the response vector. We will keep this in mind in out future analysis.

3.3.2 Normality

Normality of the errors is checked by checking the normality of the residuals. The normal qq-plot (Quantile-Quantile Plot) ?? shows quantiles of the sorted residuals averaged across time against the normal theoretical quantiles. Ideally, the relationship should be linear. Here we show the qqplot of the quantiles of residuals for the 40000th voxel, taken as a random voxel near the center of the image. We can see that the normality assumption is reasonable. In practice, this makes sense because we expect the accross-time changes/errors of estimates for a sigle voxel to be close to normally distributed.

4 Results

4.1 Behavioral analysis

We performed statistical analysis using both Python and R (The original paper use R package to fit the Logistic models). We use the library *scikit-learn* in Python and the *glm* function in *stats* in R to fit the models. Models from two library yields the same results. Following is the box plot of the behavioral loss aversion λ (median=1.94, mean=2.18, min=0.99, max=0.75). This result is consistent with that of the paper, which indicate that participants are indifferent to gambles whose gain are appoxiamtely twice as the loss.

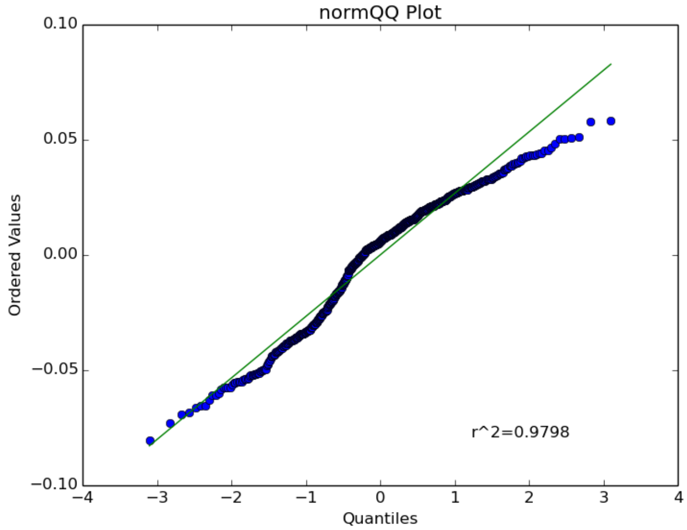
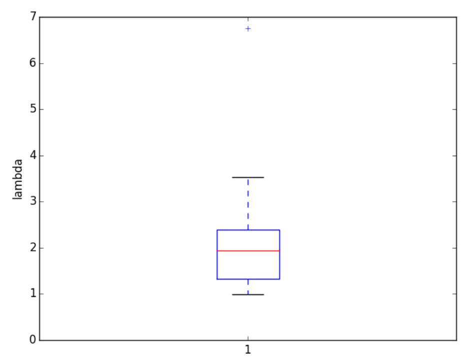


Figure 5: Normal QQ plot of the residuals for 40000 th voxel.

Figure 6: Box plot of the behavioral loss aversion λ



The accuracy of Logistic models (for 16 participants, there are 16 models in total) on the training set yielded a median of 89.78% (min=80.97%, max=99.21%). We also did the model evaluation using 10-fold cross-validation, they are still performing accuracies of a median of 89.86% (min=79.92%, max=98.45%).

4.2 Linear Regression on BOLD data

The topic we are interested in exploring is whether loss aversion reflects the engagement of distinct emotional processes when potential gains and losses are considered. In the process, we want to explore the correlation between neural and behavioral loss aversion in whole brain analysis. We also want to try to identify the regions of brain that is more activated by this loss aversion activity.

Since we want to explore the correlation between neural and behavioral loss aversion, the second step is to find out the neural loss aversion. In order to find the neural loss aversion, we perform a linear regression on the BOLD data against the parametric gain values and the parametric loss values, as explained in our model section. While implementing the linear regression, we added linear and quadratic drift in our model. These drift terms are modeling for gradual drifts across the time series.

We are especially interested in the beta coefficients of our parametric gain and parametric loss regressors, which are the first two columns in our design matrix. By looking at the coefficient values, we can get a general idea of how potential gains and potential losses affect brain activation. By plotting heat maps for gain and loss coefficients, we can identify the areas that have large coefficients; these are the areas that the brain activation is highly connected to the potential gains and losses. We choose subject

2 to plot heat maps. We plot slices 2 to 31 from the third dimension of the brain (top to bottom). The red color is associated with large coefficients and blue color is associated with small coefficients.

From the gain and loss coefficients, we can also compute the neural loss aversion. This serves the next step of looking at the correlation between neural and behavioral loss aversion. The neural gain and loss coefficients were broadly distributed and spanned zero, so it is not possible to compute the ratio of loss to gain coefficients, nor does it make much sense. Therefore, we compute the neural loss aversion at every voxel by subtracting the slope of the gain response from the (negative) slope of the loss response. Again, we plot the heat maps for subject 2s difference in gain and loss coefficients. With the neural loss aversion values calculated, we can explore how loss aversion affects brain activation when potential gains and losses are considered.

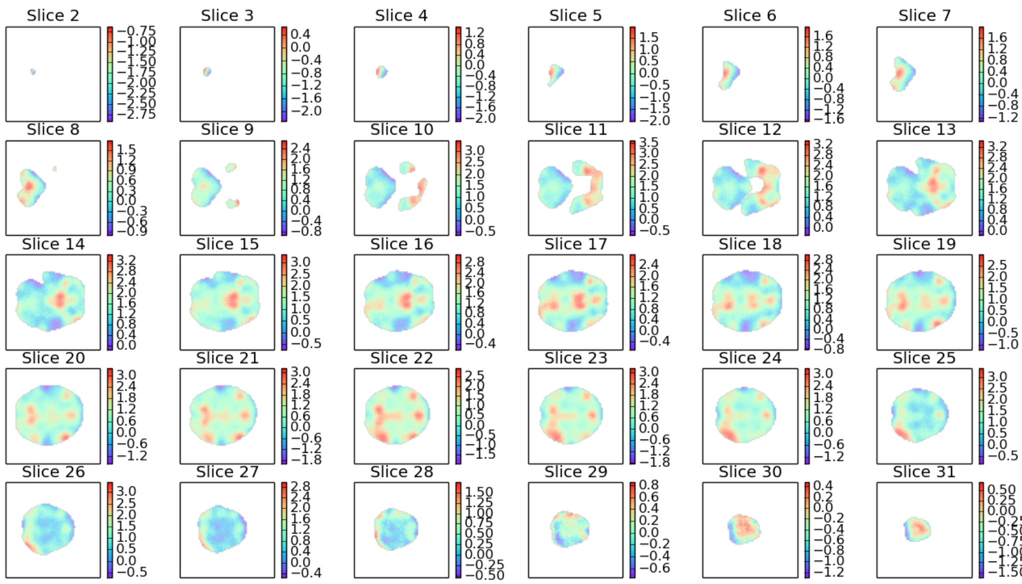


Figure 7: Coefficients of the gain values for subject 2

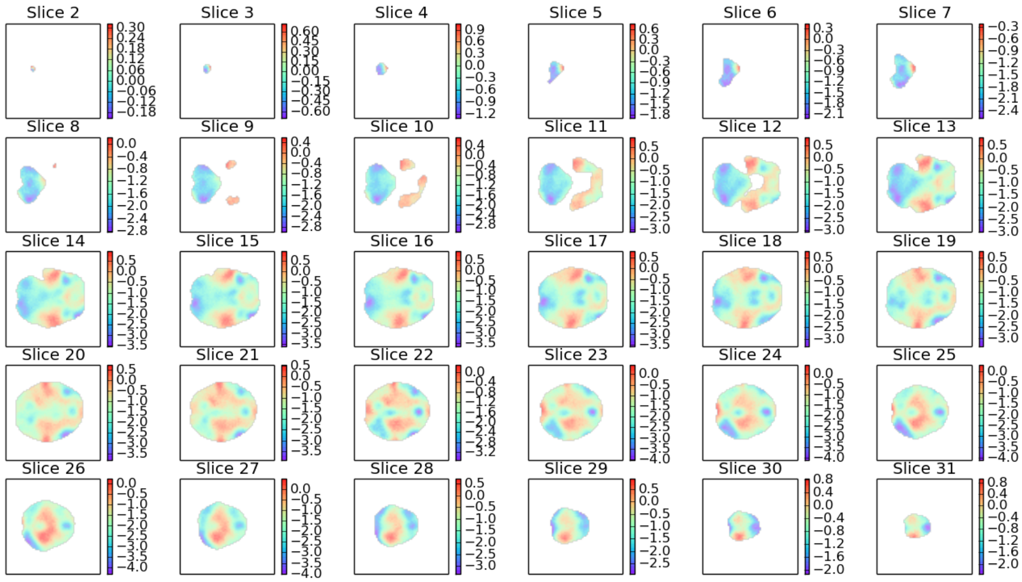


Figure 8: Coefficients of the loss values for subject 2

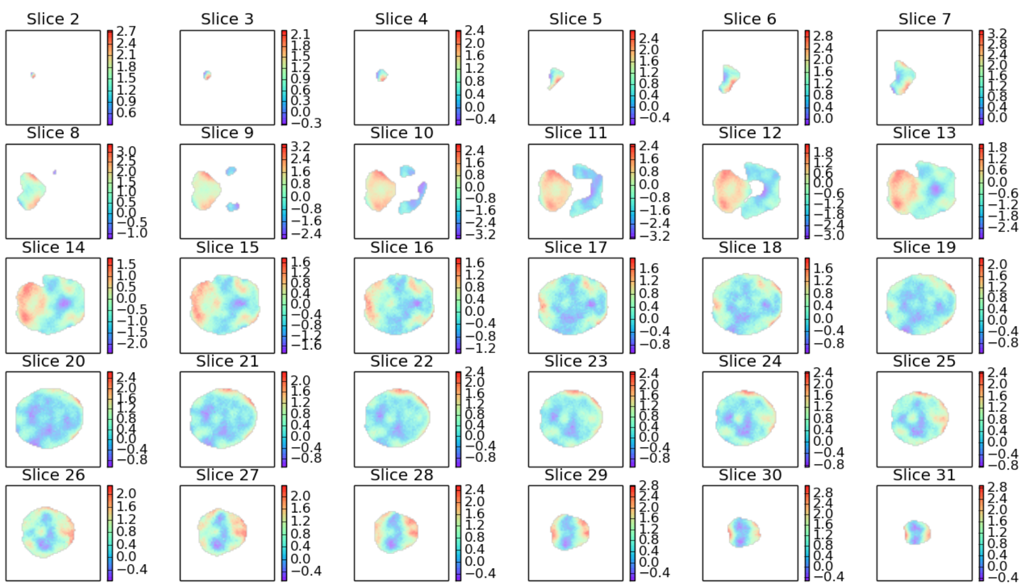


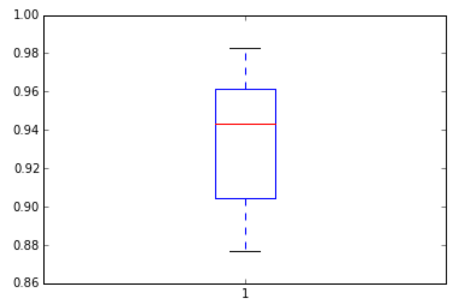
Figure 9: Difference between gain and loss coefficients for subject 2

In the near future, we will calculate the p-values and/or t statistics for each voxel of each subject and subset all the voxels with significant coefficients. Then, we can produce heat maps of the t statistics for the gain and loss coefficients of each voxel. Plotting this will show us regions with significant parametric increase in fMRI signal to increasing potential gains and regions with significant parametric decrease to increasing potential losses. On one hand, we can perform linear regression assumption check in doing so; on the other hand, we can also see the coefficients in which regions are significant and thus have significant brain activation.

4.3 Mixed-effects model on fMRI data

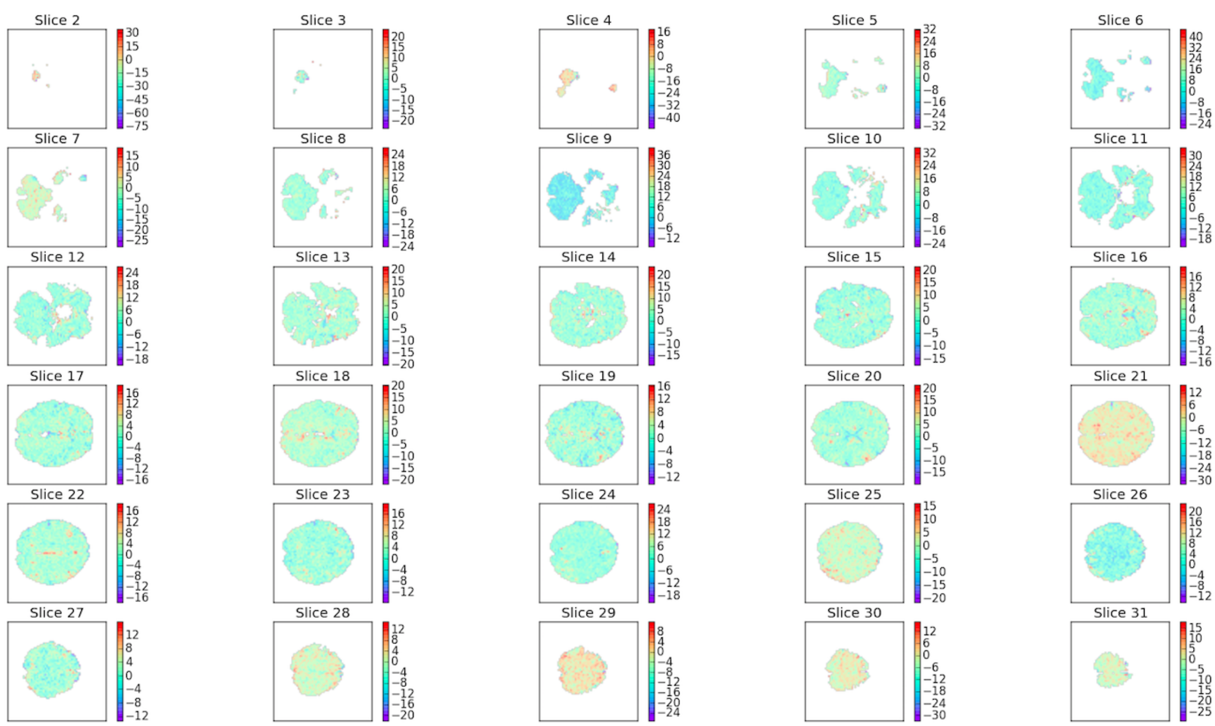
First, we did the ANOVA test for each subject each voxels, grouping by runs. The high proportion of significant ANOVA F-test (after Bonferroni correction under 0.05 significant level) shows that mixed effects model may perform well when collapsing three runs into one model.

Figure 10: Box plot of the proportion of significant ANOVA test



Following are the slices of coefficient of gain for subject 002. The mixed-effects model for each subject yielded a a median of 9.4% (min=6.4%, max=21.5%) and 8.3% (min=4.6%, max=15.4%) of proportion of significant coefficient for gain and loss separately.

Figure 11: heatmap of coefficient of gain for subject 002



5 Discussion

5.1 Issues with analyses and potential solutions

5.1.1 Selecting specific regions to further explore correlation between neural and behavioral activity

Since we have no knowledge on the sections of brain that might experience large difference in activation, it is hard for us to pick the regions to deeper explore the correspondence between neural and behavioral loss aversion.

There are two potential ways to deal with this issue. The first one is to read more paper and related articles to learn which parts of the brain are likely to react in our given scenario – faced with potential gain and loss combinations. Another way to deal with the issue to to fit a regression for every part of the brain and look for the areas with higher correspondence (higher slope). Then, we select and graph a few areas with the most significant positive or negative correlation between the parametric response to potential losses and behavioral loss aversion (ln()) across participants.

5.1.2 Producing heat map

Another issue that we are facing during our project is finding the same region to plot for each participant. We see that each region of the brain has its own standard coordinates. However, without much knowledge of fMRI, we are not sure how to use these standard coordinates to locate the regions of the brain.

From our understanding, each subject's brain is mapped onto a standard brain and we then use the coordinates for the standard brain to extract data from the areas we are interested in. However, currently, we don't have the skill to perform this step.

5.2 Further Research

We fit a linear regression model combining behavioral and BOLD data to examine the relationship of correlation between neural activity and behavioral response, we use another method which is different from what is mentioned in the paper. We add the behavioral response to the regression model on BOLD data as a predictor. We use the original 4-level response as stated below.

Moreover, if the three tests we do for the linear regression model is bad. We can plot the independents and the dependents on plots to see whether they fit a model that is different from linear regression models. There may be another reason why the performance of linear regression models are bad which is that we simplify our model that we didnt try a mixed model as the researchers in the paper did.

behavioral response	strongly accept	weakly accept	weakly reject	strongly reject
X_{behav}	1	2	3	4

And the models are following:

$$Y_i = \beta_{i,0} + \beta_{i,behav} * X_{behav} + \epsilon_i \tag{7}$$

However, since the response and level of loss and gain are potentially correlated, we might need to use stepwise regression to choose the best predictor from the regression model presented above.

5.2.1 Brain Map

We decided to not use the filtered fMRI data due to its large data size (almost 15 GB); the raw was computational less exhaustive and more managable. Yet one of the drawbacks of not using the filtered data is that we lack the brain registration to a standard anatomical template (the MNI template). Thus, much of our analysis was done by comparing subjects separately and using visuals to help motivate analysis. In the future, it may be of interest to map beta estimates from each subject onto the standard brain template to produce a a unified look at the neural loss aversion.

5.2.2 Other Potential Analysis

Looking at our data of subjects, it may be of interest to consider a demographic grouping by gender because our dataset contains the demographics of our 16 subjects, with the extra information of gender and age. A question to potentially address: Is there a significant difference to loss aversion across genders?

Additionally, it is interesting to see that the behavorial data contains a column for the response time of each gambling task. To further explore how decision are made in gambling task, we can use the response time as one of the logistic regressors. Through step-wise or criterion based model selection methods (eg. AIC and backward elimination), we can attempt to find the best regressors that influence loss aversion to most.

5.3 Discussion of Challenges

One of the major challenges is trying to make this project as reproducible as possible while following guidlines on documentation, testing functions, and attempting to produce the results of the paper using our limited understanding of fMRI data. Travis CI bugs with various versions of python, coverage failures, and errors with directory/path locations often hinder the process of smooth workflows. Collaboration between five group members is no doubt difficult as we found it hard to come up with an attainable final goal that is still rewarding.

Technically, most of us are new to python programming and reseach using git workflows, thus we have only a preliminary understanding of the various python resources available for our use. Additionally, lack of statistical understanding of some aspects of the paper has urged us to do independent research. Yet the disconnent between theory and implementation has been a major obstacle as we try to put our knowledge into practice.

Some problems can be solved or alleviated by defined checkpoints and making the effort to read and re-read the paper and ask questions. Further, as we familiarize ourselves more and more with various python modules and toolkits, results can be easier to attain and interpret.