# Explicit Categorization Ability Predictor for Biology Classification using fMRI

## Jung-Ho Byeon · Il-Sun Lee · Yong-Ju Kwon*

### Korea National University of Education

**Abstract:** Categorization is an important human function used to process different stimuli. It is also one of the most important factors affecting measurement of a person's classification ability. Explicit categorization, the representative system by which categorization ability is measured, can verbally describe the categorization rule. The purpose of this study was to develop a prediction model for categorization ability as it relates to the classification process of living organisms using fMRI.

Fifty-five participants were divided into two groups: a model generation group, comprised of twenty-seven subjects, and a model verification group, made up of twenty-eight subjects. During prediction model generation, functional connectivity was used to analyze temporal correlations between brain activation regions. A classification ability quotient (CQ) was calculated to identify the verbal categorization ability distribution of each subject. Additionally, the connectivity coefficient (CC) was calculated to quantify the functional connectivity for each subject. Hence, it was possible to generate a prediction model through regression analysis based on participants' CQ and CC values. The resultant categorization ability regression model predictor was statistically significant; however, researchers proceeded to verify its predictive ability power. In order to verify the predictive power of the developed regression model, researchers used the regression model and subjects' CC values to predict CQ values for twenty-eight subjects. Correlation between the predicted CQ values and the observed CQ values was confirmed.

Results of this study suggested that explicit categorization ability differs at the brain network level of individuals. Also, the finding suggested that differences in functional connectivity between individuals reflect differences in categorization ability. Last, researchers have provided a new method for predicting an individual's categorization ability by measuring brain activation.

**Key words:** Explicit categorization, Biology, Classification, Prediction model, fMRI

## I. Introduction

Every human classifies objects surrounding him/her. For example, all persons recognize an apple to be an "apple"; that is, it is not merely seen as an apple by one individual, all others also recognize it as an apple. All things in life are named as a result of classification. In fact, people construct a cognition system of objects and concepts related to human life through classification, which eases communication with others. The results of classification affect one's cognitive system construction, and as such classification is closely linked to human existence (Seger, 2008). Moreover, classification is also linked closely to biology, providing basic knowledge and information about nature (Margulis, 1981; Mayr, 1997). Particularly, it provides schemes for biological communication and assists with scientific knowledge generation (Margulis, 1981; Honey & Paxman, 1986). In the field of science education, classification is viewed as crucial to the understanding of complex nature (Jiang *et al.*, 2007; Seger, 2008).

Most major research on classification focuses on the understanding of classification criteria and types of classification behavior. These types of research highlight students' understanding of classification criteria, illustrate the existence of misconceptions about classification criteria, and detail characteristics of classification behavior. However, research has yet to provide a suitable

measurement method of student classification ability, which could be due to the complicatedness of the cognitive process during classification. Because some research has found that classification ability is a comprehensive and integrative cognition process that involves various other inductive cognitions (Kwon et al., 2003), it is indeed very difficult to measure an individual's classification ability through interviews, a behavior analysis and open-type paper tests. In other words, it is problematic to quantify various cognitive processes and their respective results. Additionally, the quantification of categorization triggers differences not only in the classification process but also in the ability to quantify classification ability (Jacob, 2004). Humans may respond differently to an object or event belonging to a different class or category (Ashby & Casale, 2003); that is to say, classification as a cognitive process is important for humans, basically performing according to different stimuli (Maddox et al., 2010). Classification and categorization are similar cognitive processes, but for the purpose of this paper, classification includes categorization during the cognitive process. Then, categorization becomes the most important factor in the measurement of classification ability. Most recent categorization research focuses on two different type systems known as explicit and implicit categorization (Ashby & Maddox, 2005; Ashby et al., 1998; Erickson & Kruschke, 1998; Pickering, 1997). Explicit categorization is a rule-based system tied to language function and consciousness while implicit categorization is related to non-consciousness processes (Ashby et al., 1998; Filoteo et al., 2005; Schnyer et al., 2009). It is possible to measure explicit categorization via an interview, an open-type paper test or a behavior test, but implicit categorization cannot be measured through these methods. Results of current research suggest there is a secondary output of brain activation during categorization. Neurological instruments like fMRI can aid in the detection and quantification of this brain activation during a cognitive task (Ansari & Coch, 2006). Hence, there is a need to develop a new measure of assessment for categorization ability. To meet this need, researchers developed a categorization ability prediction model for living organisms using fMRI.

## Ⅱ. Methods and Procedure

### 1. Participants

Fifty-five healthy graduate and undergraduate university students (29 male and 26 females; right-handed) were recruited for fMRI image acquisition. Participants gave informed participation consent, and the project was approved by the Ethics Committee of KNUE. Participants were divided into two groups: a model generation group comprised of twenty-seven subjects (14 males, 13 females) and a model verification group made up of twenty-eight subjects (14 males, 14 females). Researchers employed a paper task that utilized the Classification ability Quotients (CQ) to identify the classification ability distribution for each participant within the two groups (Kwon et al., 2007). The average CQ values for each group are shown in the following Table.

### 2. Procedure

#### 1) Task development for predictor generation

Table 1

*CQ distribution for the model generation and model verification groups*

| Group | Model Generation(n = 27) | Model Verification(n = 28) |
|---|---|---|
| CQ | M(SD) | M(SD) |
| | 44.99(±31.77) | 25.29(±5.07) |

The categorization process of a research task should naturally arouse the interest of each subject (Borg & Gall, 1989). The task research and development process regular seminar meetings of three science education experts-generated thirty-six tasks, each composed of nine objects, from three organism domains: animal, plant, and fungi. Eighteen of these tasks were selected for use in the study based on a calculated coincidence index of more than 90%. The study utilized two equivalent type tasks to generate a predictor of explicit categorization ability. The first one is used during MR scanning, and the other task illustrates the explicit categorization ability distribution of each subject via a CQ value. The eighteen selected tasks were randomly divided into two types of tasks, each containing 3 animal, 3 plant and 3 fungi type domains. Factor analysis was performed to evaluate construct validity. Also, to facilitate task analysis, Cronbach's alpha value and an equivalent-form reliability test were ascertained for 20 university students independent of this study.

### 2) fMRI data collection

Nine classification condition tasks: 3 animal, 3 plant, and 3 fungi, were presented during MR scanning. A blocked design was used to acquire BOLD signals from the whole brain region. fMRI scans were performed on a 3.0T MRI scanner (Forte, ISOL Technology, Korea) with a standard head coil. An EPI sequence was used for the functional imaging acquisition with the following parameters: TR / TE / $\theta$ = 3000ms / 35ms / 90°, FOV = 210mm x 210mm, matrix = 64 x 64, and slice thickness = 5mm. Axial slices parallel to the AC-PC line were acquired to provide a high-resolution image of the anatomy of the whole brain (Evans et al., 1993).

### 3) fMRI data analysis for predictor generation

Image preprocessing and statistical analysis were performed with Statistical Parametric Mapping (SPM2, Wellcome Department of Cognitive Neurology, London, UK), which is done in Matlab (Mathworks Inc. Sherborn, Mass, USA). Functional images were realigned, normalized and smoothed with an 8mm FWHM Gaussian filter (Worsley & Friston, 1995). Moreover, global scaling was added to remove the drift in BOLD signals. A general linear model was used to estimate the condition effect for individual participants (Friston et al., 1999), and the group effect was computed using one-sample t-test. The Bonferroni method makes it possible to perform multiple comparisons during statistical analysis of EPI images, but this method may cause type-Ⅱ errors (Loring et al., 2002). To obtain reliable activation data and avoid type-Ⅱ errors, a Gaussian kernel (Genovese et al., 2002) was employed. Significance level (p) was found to be 0.001, and the extent threshold was not modified (k = 0). Montreal Neurology Institute (MNI) coordinates of activation regions during explicit categorization were converted to Talairach coordinates (Talairach & Tournoux, 1988).

ROIs relevant to the extraction of time series BOLD signals were selected from a prior study on the brain network model of classification (Byeon et al., 2009). Time course BOLD signals were extracted from 9 ROIs: the left Superior Parietal Louble, the right Cuneus, the right Lingual Gyrus, the left and right Fusiform Gyrus, the left Inferior Frontal Gyrus, the left and right Middle Frontal Gyrus, the right Culmen, and the left Declive, and verified for significance using functional connectivity (Koshino et al., 2005). The functional connection strength among the various ROIs for each subject was calculated through the connectivity coefficient (CC) (Schmithorst & Holland, 2006). To generate a prediction model for explicit categorization ability, researchers employed regression analysis using the CQ and CC values obtained from each subject.

### 4) Verification of the predictor

To verify the predictive power of the developed

predictor, researchers acquired images for the twenty-eight subjects via MR scanning. Also, the CC values of each subject were calculated using time-course BOLD signals. After CC value calculation, CQ values were predicted by entering CC values into the developed prediction regression model. After calculating predicted CQ values, correlation analysis between predicted CQ and observed CQ (actual CQ) values of the twenty-eight subjects was performed.

# III. Result

## 1. Result of task analysis

For factor analysis, a KMO & Bartleet test was used to identify suitability of eighteen tasks. Suitability, as determined from factor analysis, was found to be statistical significant at $p < 0.05$. All tasks converged into one factor and the factor loading value was found to be significant with a $p$ of more than 0.50. Therefore, each

developed task is considered suitable for use as a measurement of explicit categorization ability. Also, an inter-item consistency test and an equivalent-form reliability test on the two types of tasks revealed the reliability of each task (Table 2).

## 2. Generation of predictor by fMRI data

To generate a predictor of explicit categorization ability, the functional network model from a prior study on classification was used. Statistically significant brain regions from that study's fMRI data analysis were the left Superior Parietal Louble, the right Cuneus, the right Lingual Gyrus, the left and right Fusiform Gyrus, the left Inferior Frontal Gyrus, the left and right Middle Frontal Gyrus, the right Culmen, and the left Declive (9 ROIs). More explicit representations are given in Figure 1.

To identify difference in individual brain networks, time series BOLD signals from the same

**Table 2**
*Results of the Reliability Test*

|  | f-task | p-task |
|---|---|---|
| Cronbach's alpha ($\alpha$) | 0.71 | 0.78 |
| Correlation coefficient (r) | 0.73 | |

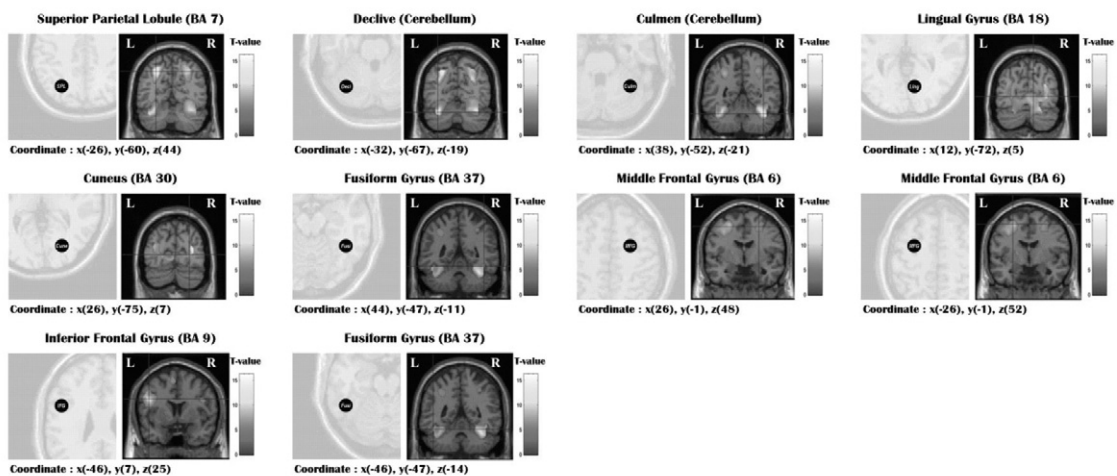\* f-task: fMRI task, p-task: CQ task



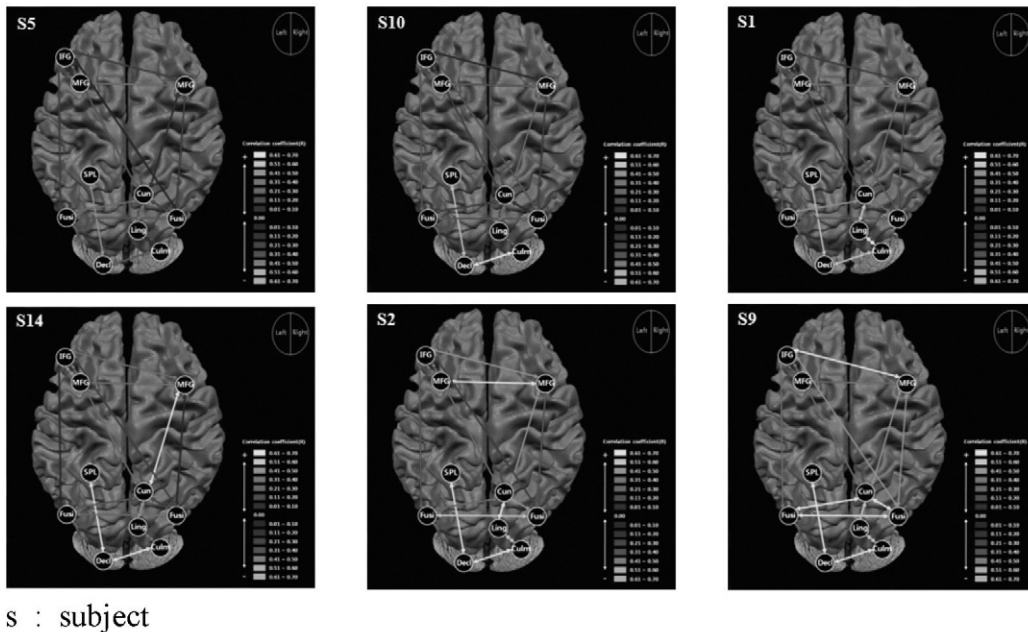**Fig. 1** *ROIs for extraction of time course BOLD signals*

s : subject

**Fig. 2** *Individual Subject Difference as shown through functional connectivity*

brain regions (9 ROIs) of each subject were extracted. Next, an individual's functional brain network was calculated using the temporal correlation coefficient (*R*) and weight (*W*) value from time series BOLD signals within ROIs (Fig. 2).

Differences in functional networks based on temporal correlation were likely the result of activated voxels as shown through cluster-based analysis, which is one type of data driven analysis (Boccalettia *et al.*, 2006; Friston *et al.*, 1993). Therefore, it is possible to calculate a connectivity coefficient (CC) value that quantifies difference in individuals' functional network (Schmithorst & Holland, 2006). Through this process, researchers quantified difference in functional networks individually and found the average CC value within groups (n = 27) to be 3.06 (±1.52).

To generate a predictor of explicit categorization ability based on individual CC values and CQ values, regression analysis was carried out with CQ and CC values (Fig. 3). The predictive power ($r^2$) of the generated predictor was found to be 0.21 by regression analysis and the correlation coefficient (*r*) was 0.46 ($p = 0.017$).

$$Y = 15.806 + 9.521X$$

Y : Predicted CQ
X : CC value

**Fig. 3** *prediction model of explicit categorization ability based on fMRI data*

### 3. Verification of predictor

To verify the predictor, the predictive power of the regression model was assessed. According to the predictor, CQ values are predicted by entering CC values into the model as independent variables. Therefore, researchers predicted CQ values from the model for each of the twenty-eight subjects. Correlations were also verified between predicted CQ values and observed CQ values. The result of correlation analysis found the relation to be statistically significant ($p < 0.05$) (Fig. 4).

Regression validity for neurological studies related to prediction models of cognitive ability regard significance to be within the 0.05 level (Chen *et al.*, 2007; Choi *et al.*, 2008; Eckert *et al.*, 2008). The predictive power of the regression
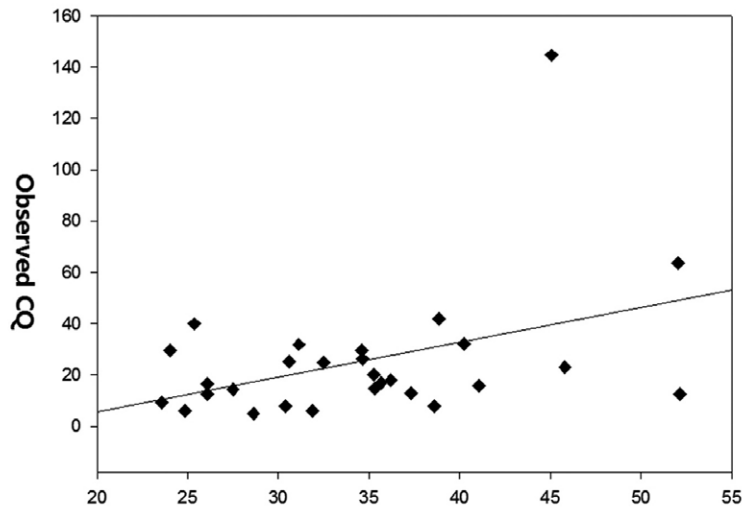
**Fig. 4** *Correlation between predicted CQ and observed CQ values*

model developed in this study is statistically significant, but the Pearson correlation is rather low. However, this study statistically verified the predictive power of the regression model through correlation analysis between predicted values and actual values. Therefore, it could be concluded that the developed predictor is a suitable prediction model for explicit categorization ability.

## IV. Discussion

Categorization ability is an important cognitive factor and essential to the investigation of classification ability (Harnad, 2005; Huang-Pollock, Maddox, & Karalunas, 2001). There are two competitive systems that explain categorization: explicit and implicit systems. According to the COVIS model, the separated two systems partially overlap and guide humans to make correct decisions. Particularly, explicit categorization is a representative measure of categorization ability because it can verbally describe the used categorization rule (Ashby *et al.*, 1998).

In this context, neurological studies suggest that the explicit categorization is related to prefrontal activation for top-down and bottom-up type decision making. The result of this study

is similar to these suggestions because the brain connectivity in the prefrontal region differed according to the categorization ability of a subject. According to the result of this study, it could be said that explicit categorization ability can be identified at the brain level of an individual. Second, this study suggests that differences in functional connectivity between individuals reflect differences in categorization ability. Third, having been verified with a group of participants independent of this study, researchers offer the developed regression model as a predictor of explicit categorization ability through brain network measurement. Lastly, it could be said that prediction of explicit categorization ability is possible by assessing brain connectivity.

The result of this study may explain several mechanisms related to differences in individuals. Also, the research method of this study led to the development of a new measure of assessment for cognitive ability. Certainly, this paper was limited in subject number, yet the result of this study can provide more objective data about individual's explicit categorization. It, however, cannot predict causes of categorization ability differences among individuals. Further studies on a different and larger scale are needed

before applying the developed prediction model beyond this study. In addition, it is necessary to investigate causes of categorization ability differences like categorization strategies.

# References

Ansari, D. & Coch, D. (2006). Bridge over troubled waters: education and cognitive neuroscience. Trends in Cognitive Sciences, 10(4), 146-151.

Ashby, F. G. & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. Attention and implicit learning, Amsterdam: John Benjamins Publishing Company, 109-141.

Ashby, F. G. & Maddox, W. T. (2005). Human category learning. Annual Review of Psychology, 56, 149-178.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. Psychological Review, 105, 442-481.

Boccalettia, S., Latorab, V., Morenod, Y., Chavezf, M., & Hwang, D. U. (2006). Complex networks: structure and dynamics. Physic Reports. 424(4-5), 175-308.

Borg, W. R. & Gall, M. D. (1989). Educational research: An introduction, 5th Ed. New York, NY: Longman Inc.

Byeon, J. H., Lee, J. K., & Kwon, Y. J. (2009). Brain activation pattern and functional connectivity network during classification on the living. Journal of Korean Association for Research in Science Education, 29(7), 751-758.

Chen, C., Xue, G., Dong, Q., Jin, Z., Li, T., Xue, F., Zho, L., & Guo, Y. (2007). Sex determines the neurofunctional predictors of visual word learning. Neuropsychologia, 45, 741-747.

Choi, Y. Y., Shamosh, N. A., Cho, S. H., Deyoung, C. G., Lee, M. J., Lee, J., Kim, S. I., Cho, Z., Kim, K., Gray, J. R., & Lee, K. H. (2008). Multiple bases of human intelligence revealed by cortical thickness and neural activation. The Journal of Neuroscience, 28(41), 10323-10329.

Eckert, M. A., Lombardino, L. J., Walczak, A. R., Bonihla, L., Leonard, C. M., & Binder, J. R. (2008). Manual and automated measures of superior temporal gyrus asymmetry: Concordant structural predictors of verbal ability in children. NeuroImage, 41, 813-822.

Erickson, M. A. & Kruschke, J. K. (1998). Rules and exemplars in category learning. Journal of Experimental Psychology: General, 127, 107-140.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993). 3D Statistical neuroanatomical model from 305 MRI volumes. IEEE Conference Record, Nuclear Science Symposium and Medical Imaging Conference (San Francisco). P 1813-1817.

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., & Matthews, S. (2005). Cortical and subcortical brain regions involved in rule-based category learning. Neuroreport, 16(2), 111-115.

Friston, K. J., Frith, C. D., & Frackowiak, R. S. J. (1993). Time-dependent changes in effective connectivity measured with PET. Human Brain Mapping, 1, 69-79.

Friston, K. J., Holmes, A. P., & Worsley, K. J. (1999). How many subjects constitute a study? NeuroImage, 10, 1-5.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. NeuroImage, 15, 870-878.

Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. In C. Lefebvre & H. Cohen (Eds.), Handbook on Categorization, Elservier.

Honey, J. N. & Paxman, H. M. (1986). The importance of taxonomy in biological education at advanced level. Journal of Biological Education, 20, 103-111.

Huang-Pollock, C. L., Maddox, W. T., & Karalunas, S. L. (2011). Development of implicit and explicit category learning. Journal of

Experimental Child Psychology, 109, 321-335.

Jacob, E. K. (2004). Classification and Categorization: A difference that makes a difference. Library Trends, 52(3), 515-540.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J. & Riesenhuber, M. (2007). Categorization training results in shape and category-selective human neural plasticity. Neuron, 53, 891-903.

Koshino, H., Carpenter, P. A., Minshew, N. J., Cherkassky, V. L., Keller, T. A., & Just, M. A. (2005). Functional connectivity in an fMRI working memory task in high-functioning autism Neuroimage 24, 810-21.

Kwon, Y. J., Choi, S. J., Park, Y. B. & Jeong, J. S. (2003). Scientific thinking types and processes generated in inductive inquiry by college students. Journal of Korean Association for Research in Science Education, 23(3), 286-298.

Kwon, Y. J., Lee, J. K., & Lee, I. S. (2007). Development of the classification ability quotient equation through the analysis of science teachers' classification knowledge generated in the pollen classification task. The Secondary Education Research, 55(3), 21-43.

Loring, D. W., Meador, K. J., Allison, J. D., Pillai, J. J., Lavin, T., Lee, G. P., Balan, A., & Dave, V. (2002). Now you see it, now you don't: statistical and methodological considerations in fMRI. Epilepsy & Behavior, 3, 539-547.

Maddox, W. T., Pacheco, J., Reeves, M., Zhu, B., & Schnyer, D. M. (2010). Neuropsychologia, 48, 2998-3008.

Margulis, L. (1981). How many kingdoms? Current views of biological classification. The American Biology Teacher, 43, 482-489.

Mayr, E. (1997). This is biology: the science of the living world. Cambridge. Massachusetts: The belknap press of Harvard University press.

Pickering, A. D. (1997). New approaches to the study of amnesic patients: What can a neurofunctional philosophy and neural network methods offer? Memory, 5, 255-300.

Schmithorst, V. J. & Holland, S. K. (2006). Functional MRI evidence for disparate developmental processes underlying intelligence in boys and girls. NeuroImage, 31, 1366-1379.

Schnyer, D. M., Maddox, W. T., Ell, S., Davis, S., Pacheco, J., & Verfaellie, M. (2009). Prefrontal contributions to rule-based and information-integration category learning. Neuropsychologia, 47(13), 2995-3006.

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. Neuroscience and Biobehavioral Review, 32, 265-278.

Talairach, J. & Tournoux, P. (1988). Co-Planner stereotaxic atlas of the human brain. New York: Thieme Medical publisher, Inc.

Worsley, K. J. & Friston, K. J. (1995). Analysis of fMRI time-series revisited-again. NeuroImage 2, 173-181.