



# Project no: 027787

# DIRAC

## **Detection and Identification of Rare Audio-visual Cues**

Integrated Project IST - Priority 2

## **DELIVERABLE NO: D4.9** Neuronal basis of categorization

Date of deliverable:2008-12-31 Actual submission date:2009-01-30

Start date of project: 01.01.2006

Duration: 60 months

Organization name of lead contractor for this deliverable: Leibniz Institute for Neurobiology (LIN)

Project co-funded by the European Commission within the Sixth Framework Program			
(2002-2006)			
Dissemination Level			
PU	Public		
PP	Restricted to other program participants (including the Commission Services)		
RE	Restricted to a group specified by the consortium (including the Commission	Х	
	Services)		
CO	Confidential, only for members of the consortium (including the Commission		
	Services)		





# D4.9 NEURONAL BASIS OF CATEGORIZATION Leibniz Institute for Neurobiology (LIN) The Hebrew University of Jerusalem (HUJI)

#### Abstract:

It has been suggested that different sub-modules in the striatum play different roles in different learning scenarios. Specifically, it is claimed that the dorsal striatum (the head of the putamen and caudate nucleus) is involved in learning simple rule-based category structures while the ventral striatum is more strongly involved in learning complex category structures. In the current study we suggest that this functional definition of the striatum in category learning is insufficient. Using fMRI (functional Magnetic Resonance Imaging), we show that the level of neuronal activation in the dorsal striatum is strongly associated with the type of the information provided during category learning. Particularly, when learning a complex rule-based categorization principle by a comparison of only few objects noted to belong to different categories (different-class exemplars), the level of neuronal activation in the dorsal striatum is significantly higher than when learning the same rule comparing objects noted to be from the same category (same-class exemplars) or when the task does not require learning a new categorization rule. Interestingly this is the case although the objective information quantity is identical in the two category learning conditions. We suggested that different neuronal circuitries evolved in order to process different information building blocks according to their expected information content or expected usability, rather than their actual usability or the-to-be learned categorization principle and its complexity.





## **Table of Content**

Abstract	1
Table of content	2
Introduction	3
Methods	4
Results	10
Conclusion	14
References	15
	Abstract Table of content Introduction Methods Results Conclusion References





## 1. Introduction

Studies that tested the neuronal mechanisms of *category learning* suggest that different nuclei of the basal-ganglia play a major part in this process (e.g. Ashby et al, 2007; Nomura et al., 2007; O'Doherty, 2008; Turk-Browne, 2009). More specifically, it has been suggested that category learning involves both implicit (procedural) and explicit (declarative) learning processes, each associated with different nuclei within the basal ganglia (as well as with other brain areas). Furthermore, it has been suggested that the engagement of either one of these learning systems depends on the structure of the learned categories (Ashby & Maddox, 2005; Ashby et al., 2007) or the stages and advances of learning (Seger & Cincotta, 2005).

In the current study we present an alternative viewpoint for the previously reported disparate systems for category learning. We claim that fundamentally, category learning involve the computation of both within category variation and between category variation as these are indicated to the learner when comparing objects identified to belong to the same category (Same-Class Exemplars) or to different categories (Different-Class Exemplars). Recent findings (Hammer et al., 2007, 2008, 2009a, 2009b) suggest that learning from same-class exemplars comparison is an automated process by which people can intuitively learn simplified categorization principles, whereas learning from different-class exemplars comparison requires explicit reasoning that potentially may end up with an enhanced performance. Accordingly we suggest that different systems for category learning are not necessarily associated with the structure of the learned categories, but rather with the nature of the information provided to the learner and the way this information can be used. That is, processing same-class indications is likely to be associated with one neuronal mechanism while processing different-class indications is likely to be associated with another (irrespectively to the learned category structure).

In order to test our hypothesis and confront them with alternative previous findings concerning the neural correlate of rule-based category learning we undertook the following methodological precautions: (1) We disassociated the two learning conditions so that in one condition participants were trained with only same-class indications while in the other they were trained with only different-class indications. We further compared these two category learning conditions with a control condition in which no category learning took place, but in which participates were provided with similar visual input and they were required to produce similar responses (motor output). (2) In each one of the two category learning conditions, the learning phase was disassociated from the test phase. Furthermore, participants were not provided with any feedback for their performance during the task. (3) We selected the same-





class and different-class indications so that in the two category learning conditions the objective information quantity provided for the participants was identical. (4) In the two category-learning conditions participants were always required to learn a *complex* rule-based categorization principle in which the conjunction of two out of four task relevant feature-dimensions had to be identified. Objectively, such categorization principle is equally likely to be learned by either the same-class or the different-class indications that were provided to the participants during the learning phase.

### 2. Methods

#### 2.1. Participants

Fourteen participants, 4 males and 10 females, with an average age of 28.9 (age range = 22-41 years old), with normal or corrected to normal vision participated in the experiment. Participants gave written informed consent to the study, which was approved by the ethics committee of the University of Magdeburg.

#### 2.2. Materials

Twelve sets of computer generated grayscale images of "creature-like" objects were used as stimuli. Each creature set was characterized by four binary feature-dimensions that potentially could determine creature categories within the given set. Participants were required to learn the specific categorization rule for each set separately. This categorization rule required participants to identify the conjunction of two task relevant feature-dimensions out of four. In addition to the creature-like object, each stimulus contained low-contrast background pattern with either blurred circles or blurred squares scattered in a varying pattern. These background shapes were used for the control task. Fig 1.a illustrate few stimuli from one stimuli set.







Figure 1. (a) Example from one stimulus set. Each stimulus set vary in four feature-dimensions. (b) The hypotheses table illustrating all the possible combinations for relevant ("1") or irrelevant ("0") feature-dimensions. (c) Same-Class pairs indicating d3 and d4 as irrelevant. (d) Different-Class pairs indicating d2 and d1 as relevant.

#### **2.3. Behavioral procedure**

In a training session conducted shortly before the fMRI test session, each participant has become familiarized with the stimuli that were later used in the scanner. Participants were asked to perform a simple same/different category decision (by pressing a mouse key) with reference to pairs of creatures simultaneously presented on a computer monitor. Participants *did not* receive any feedback for their decision, so that at this point they were not experienced with the to-be learned categorization rule. The motivation to engage participants in this unsupervised discrimination task was only to familiarize participants with the stimulus sets and their different features. This allowed participant to map the potentially informative variations within each stimulus set.

Mapping the variations in the four potentially task relevant dimensions, for each stimulus set, enables participants to consider 16 different possibilities for categorizing the creatures as these described in the Hypotheses Table in Fig 1.b: On one extreme, Hypothesis 1 (H1) represent a scenario in which neither one of the dimensions is relevant for categorization. On the other extreme, H16 represent a scenario in which all the dimensions are relevant for categorization. In between these two alternatives, there are all the other combinations for relevant and irrelevant dimensions. It is important to note that participants *were not informed* about the number of varying dimensions or the number of task relevant dimensions needed to be identified later in the category learning task.





Shortly after concluding the training session outside the magnet, participants started the fMRI session in which they were tested in three conditions: Category learning from *Same-Class* exemplars comparison, category learning from *Different-Class* exemplars comparison and a *Control* condition in which participants did not have to learn any categorization principle. Note that in each one of the category learning blocks a different stimuli-set was used so that each block required learning a new categorization rule. Each participant was tested on 6 different category learning tasks in the Same-Class condition, 6 tasks in the Different-Class condition, as he also performed 6 tasks in the Control condition. The blocks from the three conditions were intermixed.

Each block in the two category-learning conditions started with a learning phase. In the two learning phase, same-class and different-class indications were selected so to provide the same amount of information: In the Same-Class condition participants were asked to learn the categorization principle when being trained with only two same-class indications as illustrated in Fig 1.c. In each pair two creatures were presented together with the sign "=" between the two, providing an indication for the participant that these two creatures are of the same kind. From the first sameclass pair participants could learn that two creatures from the same kind may differ in their tails (marked as d3 in Fig 1.a). That is, tail cannot be relevant for categorization since the within category variation in this feature is as large as the total variation. From the second same-class pair the participant could learn that d4 (the gap between the eyes) is also not relevant for categorization. Each one of these two indications provided participants with exactly 1 bit of information since it eliminates half of the remaining possible hypotheses illustrated in Fig 1.b. Together, the two indications provide 2 bits leaving only the four possibilities in which d3 and d4 are both irrelevant (marked by "0"). This leaves participants only with the possibility that H1, H5, H9 or H13 are valid hypotheses.

In the *Different-Class* condition participants were asked to learn the categorization principle when being trained with only two different-class indications (Fig 1.d). In each pair, two creatures were presented together with the sign " $\neq$ " between the two, indicating that these two creatures are from different kinds. From the first different-class pair the participant could learn that when two creatures differ only in the spikes on their back (marked as d2 in Fig 1.a), it is sufficient for knowing that the two are not from the same kind. That is, d2 is relevant for categorization. From the second different-class indication the participant could similarly learn that d1 is also relevant for categorization. Each one of these two indications also provided the participants with 1 bit of information by eliminating half of the remaining hypotheses. Together, the two indications provide 2 bits leaving only the four possibilities in





which d1 and d2 are both relevant (marked by "1"). This leaves participants only with the possibility that H13, H14, H15 or H16 are valid hypotheses.

In each block, immediately after being trained with either same-class or different-class indications, participants were tested on the categorization rule they had just learned in the learning phase that started the block. The test phase included 8 test trails. In each trial a pair of creatures was presented with no indication for the categorical relation between the two creatures. The two creatures presented together always differed in two out of four features. Participants had to decide whether the two creatures are from the same category or not according to what they have just learned. Only when the two creatures were identical in the two task relevant feature, participants had to decide "same". Participants did not receive any feedback for their decisions. After the test phase ended, the block was terminated.



Figure 2. (A) An illustration for one block in the same-class condition. The block starts with an indication for participant identifying the block type (same/different/shape; duration = 4 seconds). It was then followed by a learning phase (duration = 12 seconds) in which four same-class indications were presented. After the learning phase was completed, the participant started the test phase (duration = 24 seconds) in which he had to decide whether the presented paired creatures are of the same kind or different-kind. (B) An illustration for one block in the different-class condition. (C) An illustration for one block in the control condition. In this condition, during the "Learning" phase, no category learning took place but rather participants had to attend the low contrast background shapes.

The general structure, stimuli type and the duration of a block in the *Control* condition were identical to those in the category learning blocks. Nevertheless here participants were not required to learn to categorize the presented creatures but rather they were simply asked to decide whether the low-contrast background shapes in the two stimuli are same (square-square/circles-circles) or different (square-circle). These low-contrast pattern shapes were also part of the stimuli used in the category learning conditions, but in these conditions they had to be ignored. Using different distribution patterns of the background shapes (see some variations in Fig 1.a), low-contrast,





blurring and noise, made this task perceptually demanding and required participants to be attentive to the stimuli when performing the task. This control conditions enabled us to discriminate the neuronal activation associated with category learning from the neuronal activation associated with the visual input, attention and motor responses (eye saccades and key pressing) when analyzing the BOLD signal. Fig 2 provides a description for block structure in each one of the three conditions.

#### 2.4. Scanning procedure

In the fMRI Session, stimuli were displayed using the Presentation software (http://:www.neurobs.com) run on an Intel® based PC computer and back-projected onto a screen which could be viewed via a mirror mounted on the head coil. The distance between the subject's eyes and the screen was 59 cm. The screen was 325x260 mm which is appropriate for an angle of +-15°. In each experimental trial, a pair of stimuli was simultaneously presented in the center of the computer screen. Each stimulus occupied  $10 \times 10$  cm on the screen, and the two stimuli were separated by a gap of 6.5 cm. Participants responded directly using the two keys of a respond box.

The measurements were carried out on a 3 Tesla scanner (Siemens Trio, Erlangen, Germany) equipped with an eight channel head coil. A 3D anatomical data set of the subject's brain (192 slices of 1 mm each) was obtained before the functional measurement. Additionally, an *Inversion-Recovery-Echo-Planar-Imaging* (IR-EPI) with the identical geometry as in the functional measurement was acquired. For fMRI 550 functional volumes were acquired in 18:20 min using an echo planar imaging (EPI) sequence (echo time (TE), 30 ms; repetition time (TR), 2000 ms; flip angel, 80°; matrix size, 64 x 64; field of view, 19.2 cm x 19.2 cm; 33 slices of 3 mm thickness with 0.3 mm gaps). During scanning, the subjects wore earplugs and their head was fixed with a cushion.

#### 2.5. fMRI data analysis

The functional data were analyzed with the software BrainVoyager<sup>TM</sup>QX 1.8.6 (Brain Innovation, Maastricht, The Netherlands). A standard sequence of preprocessing steps, 3D-motion correction, linear trend removal, and filtering with a high-pass of three cycles per scan was performed. The functional data set was projected to the IR-EPI-images, co-registered with the 3D-data set, and then transformed to Talairach-space.

For the fixed-effects GLM-analysis we defined the following predictors for each 1 minute block: Indicate (0-3.5 s), Learning1 (3.5-9.5 s), Learning2 (9.5-15.5 s),





Test (15.5-39.0 s), Rest 39.0-60 s). These were convolved with the two-gamma hemodynamic response function using the default parameters implemented in BrainVoyager<sup>TM</sup>QX. First we calculated a balanced contrast using all predictors of the *Different-Class* condition and the *Same-Class* condition vs. all predictors of the *Control* condition. This was done to produce a mask for further analysis (see below) which only includes brain regions which show stronger activation in the *Different-Class* condition compared to the *Control* condition. Within this mask, we then calculated further contrasts to reveal specific effects for the *Different-Class* vs. *Same-Class* condition during the two encoding phases and the recall phase by using a significance level of p < 0.02 (Random effect analysis).

#### 2.6. Behavioral findings analysis

We measured participant ability to learn the new categories by using the nonparametric sensitivity measure A' (Grier, 1971), calculated from participant Hits (correctly identifying two creatures as belonging to the same category) and False-Alarms (incorrectly identifying two creatures as belonging to the same category). A' = 0.5 represents chance performance, A' = 1 represents perfect performance, and 0 < A' < 0.5 represents response confusion. For each participant we calculated his or her average performance in each condition separately (the average performance in the six blocks from each condition).

#### 3. Results

One sample t-tests showed that participant performances in all three conditions were better than chance (A' = 0.5): Control (background shapes) condition (mean A' = 0.96), t(13) = 50.21, p < 0.001. Same-class indications condition (mean A' = 0.64; SD = 0.17), t(13) = 3.10, p < 0.01. Different-class indications condition (mean A' = 0.84; SD = 0.17), t(13) = 7.23, p < 0.001. In addition, there was significantly better performance in the different-class condition than in the same-class condition, t(13) = 3.52, p < 0.005. These results are illustrated in Fig 3.







Figure 3. Left – each participant is plotted twice on a ROC diagram, once for the same-class condition and once for the different-class condition. Right – mean A' prime in the three experimental conditions.

Figure 4 presents the neural correlate of the above behavioral findings. A random effect analysis for all 14 participants shows a significantly higher activation in the right dorsal striatum, during the learning phase, when participants were trained with different-class indications (contrasted with the same-class indications condition, using random effect analysis with threshold of p < 0.02).







Figure 4: Bottom left – time course of the neuronal activation in the right dorsal striatum (averaged for all participants for all blocks within each specific condition). Brain slices show the differences in the neural activation taking place in the right dorsal striatum during the learning phase.

One possible way to interpreted the above findings is that some participants performed the category learning task in the same-class condition at near chance level (see the ROC diagram in Fig 3), but almost all participants had performed very well in the different-class condition. Therefore, it might be that the level of neuronal activation in the dorsal striatum during the learning phase is correlated with the expected behavioral performance during the following test phase. That is, the above findings may suggest that the differences between the two category learning conditions are at most quantitative, but they are not associated with the differences in the quality of the provided information.

In order to confront the possibility that the observed neural correlate represent only the quantitative difference in the usability of the two comparison types (sameclass indications were less usable for the subjects than different-class exemplars indications), we analyzed the neuronal correlate of performances for the upper median of the participants (according to participants' performance in the same-class condition). Here we find again a significantly better performance in the different-class condition (mean A' = 0.92; SD = 0.03) than in the same-class condition (mean A' = 0.76; SD = 0.06), t(6) = 5.20, p < 0.01. Nevertheless, now the mean performancedifference between the two category learning tasks have been narrowed: Mean difference (different-class condition – same-class condition) when testing all 14 subject is A'-diff = 0.20 (SD = 0.21) while the mean difference for the best performers is A'-diff = 0.15 (SD = 0.08). These results are illustrated in Fig 5.







Figure 5: Left – each one of the best seven participants is plotted twice on a ROC diagram, once for the same-class condition and once for the different-class condition. Right – mean A' prime of the seven best participants in the three experimental conditions.

Figure 6 present the neural correlate of the behavioral findings for the best seven participants. Here we see again a significantly higher activation in the right (as well as the left) dorsal striatum, during the learning phase, when participants were trained with different-class indications (contrasted with the same-class condition, when using again the threshold of p < 0.02).



Figure 6. Bottom left – time course of the neuronal activation in the right dorsal striatum (averaged for the seven best participants for all blocks within each condition). Brain slices show the differences in the neural activation taking place in the right (and left) dorsal striatum during the learning phase when using the same threshold (p < 0.02) used for Fig A2.4. Right – correlations between the neuronal activation in the right dorsal striatum during the learning phase with participants (all 14) performance level (in the test phase) for each category learning task separately.

We further analyzed the correlations between the neuronal activation in the right dorsal striatum during the learning phase with participants' (all 14) performance level for each category learning task separately. This analysis shows that there is no correlation between participants' performance and the level of the neuronal activation within each one of the two category learning conditions: Same-class condition r(14) = -0.21, p > 0.90; different-class condition r(14) = 0.20, p > 0.50.

Taken together, the above analyses suggest that the level of neuronal activation in the dorsal striatum is not directly associated with participants' level of





performance but rather it is mainly associated with processing informative differentclass indications (i.e. informative between category variation).

#### 4. Conclusions

The above findings suggest that the human brain evolved so that some neuronal circuitries are engaged mainly when learning requires processing informative different-class indications. We showed here that the dorsal striatum is significantly more associated with processing informative different-class indications than it is associated with processing similarly informative same-class indications. In fact, neuronal activation in the dorsal striatum when processing same-class indications was no higher than when participants performed a control task that involved processing similar visual input and producing the same motor responses but did not require learning new categorization principles. We suggest that this result from the fact that in most everyday life scenarios, same-class indications act as a more useful source of information, usable in most category learning scenarios, even if only for a limited extent. On the other hand, different-class indications which only rarely can much the information quantity of same-class indications may have great value when learning more specific category structures or for finalizing the learning process (Hammer et al., 2008).

The functional imaging findings presented here present a challenge for previous findings concerning the possible role of the dorsal striatum in category learning. Specifically for those claiming that the dorsal striatum is associated with category learning whenever the task involved learning a simple categorization rule (e.g. Ashby et al., 2007) or for finalizing the learning process (e.g. Seger & Cincotta, 2005). Here we showed that the dorsal striatum is significantly more involved in category learning (of a complex rule) when learning involves processing different-class indications, but not when learning involves processing same-class indication with no dependency on the learned category structure or the stage of learning it. Alternatively, we suggest that the dorsal striatum is more involved in processing informative between-category variation irrespectively of the structure of the learned categories, the complexity of the categorization rule or the stage of learning. This implies that the differential neuronal activity in the dorsal striatum, as previously documented but others, mainly result from different requirements in processing the between category variation when performing different types of supervised category learning tasks.





## **5. References**

- Ashby F. G., Ennis J. M., & Spiering B. J. (2007). A neurological theory of automaticity in perceptual categorization. *Psychological Review*, 114(3), 632-656.
- Ashby, F.G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149-178.
- Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., & Hochstein, S. (2008). Comparison processes in category learning: From theory to behavior. *Brain Research*. 1225, 102-118.
- Hammer, R., Diesendruck, G., Weinshall, D., & Hochstein, S., (2009b, accepted). The development of category learning strategies: What makes the difference? *Cognition*.
- Hammer, R., Hertz, T., Hochstein, S., & Weinshall, D. (2009a, in press). Category learning from equivalence constraints. *Cognitive Processing*.
- Hammer, R., Hertz, T., Hochstein, S., & Weinshall, D. (2007). Classification with positive and negative equivalence constraints: Theory, computation and human experiments. In F. Mele, G. Ramella, S. Santillo, & F. Ventriglia (Eds), *Brain, Vision, and Artificial Intelligence: Second International Symposium, BVAI* 2007. Lecture Notes in Computer Science (pp. 264-276). Berlin Heidelberg: Springer-Verlag Press.
- Nomura E. M., Maddox W. T., Filoteo J. V., Ing A. D., Gitelman D. R., Parrish T. B., Mesulam M., & Reber P. J. (2007). Neural Correlates of Rule-Based and Information-Integration Visual Category Learning. *Cerebral Cortex*, 17(1), 37-43.
- O'Doherty J. P. (2008). Lights, Camembert, Action! The Role of Human Orbitofrontal Cortex in Encoding Stimuli, Rewards, and Choices. *Annals of the New York Academy of Sciences*, 1121, 254-272.
- Seger, C. A., & Concotta, C.M. (2005). The roles of the caudate nucleus in human classification learning, *J. Neurosci.* 25, 2941–2951.
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009, in press). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*.