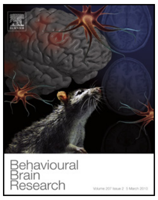
Behavioural Brain Research 311 (2016) 239–246

Contents lists available at ScienceDirect

Behavioural Brain Research

journal homepage: www.elsevier.com/locate/bbr



Research report

An interplay of fusiform gyrus and hippocampus enables prototype- and exemplar-based category learning

Robert K. Lech[a,c], Onur Güntürkün[b,c], Boris Suchan[a,c,*]

[a] Institute of Cognitive Neuroscience, Department of Neuropsychology, Ruhr University Bochum, Germany
[b] Institute of Cognitive Neuroscience, Department of Biopsychology, Ruhr University Bochum, Germany
[c] International Graduate School of Neuroscience, Ruhr University Bochum, Germany

HIGHLIGHTS

- Different brain structures for prototype- and exemplar-based category learning has been proposed.
- Exemplar based category learning is associated with fusiform gyrus activation.
- Exception learning is associated with hippocampus activation.
- Coupling between Hippocampus and fusiform gyrus activation showed a time displaced course for categorization of Prototypes and Exceptions.

ARTICLE INFO

Article history:
Received 27 January 2016
Received in revised form 19 May 2016
Accepted 23 May 2016
Available online 24 May 2016

Keywords:
Category learning
Exemplar-based learning
Prototype-based learning
Hippocampus
Fusiform gyrus

ABSTRACT

The aim of the present study was to examine the contributions of different brain structures to prototype- and exemplar-based category learning using functional magnetic resonance imaging (fMRI). Twenty-eight subjects performed a categorization task in which they had to assign prototypes and exceptions to two different families. This test procedure usually produces different learning curves for prototype and exception stimuli. Our behavioral data replicated these previous findings by showing an initially superior performance for prototypes and typical stimuli and a switch from a prototype-based to an exemplar-based categorization for exceptions in the later learning phases. Since performance varied, we divided participants into learners and non-learners. Analysis of the functional imaging data revealed that the interaction of group (learners vs. non-learners) and block (Block 5 vs. Block 1) yielded an activation of the left fusiform gyrus for the processing of prototypes, and an activation of the right hippocampus for exceptions after learning the categories. Thus, successful prototype- and exemplar-based category learning is associated with activations of complementary neural substrates that constitute object-based processes of the ventral visual stream and their interaction with unique-cue representations, possibly based on sparse coding within the hippocampus.

1. Introduction

Every day we are confronted with a bewildering variety of objects. Since we are unable to learn about each object separately, we deal with them at the category level. Categorization is the ability to generalize various stimuli into a single class, to extrapolate the categorical knowledge to new members of the stimulus classes, and to discriminate between different classes [1]. The ability to categorize effectively reduces information load and enables us to cope with the constantly changing environment [2]. Grouping similar objects reduces computational demands and enables an organism to use its resources for other purposes [3].

Research focusing on category learning has put forward various differing computational models in order to explain categorization processes for a review, see Ashby et al. [1]. Some of them are exemplar-based and thus assume storage of individual instances of one category and subsequent generalization when being faced with a new stimulus of the same category [4,5]. When being faced with a novel stimulus, its belonging to a certain category is determined by a comparison to previously encountered stimuli. One example would be a neurologist who tries to determine whether or not he sees a brain tumor on an MR image. If it would be similar to pre-

[*] Corresponding author at: Institute of Cognitive Neuroscience, Department of Neuropsychology, Ruhr University Bochum, Universitätsstraße 150, D-44780 Bochum, Germany.
E-mail address: Boris.Suchan@rub.de (B. Suchan).

http://dx.doi.org/10.1016/j.bbr.2016.05.049
0166-4328/© 2016 Elsevier B.V. All rights reserved.

viously encountered and memorized MR images of brain tumors, he would classify it accordingly. On the contrary, prototype-based models are developed from an abstracting of the central tendencies of stimuli from one category [3]. The neurologist from the previous example would not rely on the memorization of every single tumor image, but instead he would have condensed the previous images into a summary representation to which he then would compare new MR images. Furthermore, there are hybrid models like the cluster-based SUSTAIN model [6]. SUSTAIN explains category learning by initially assuming a very simple category structure that is represented by a single cluster that codes features and four categories. Surprising events, e.g. stimuli that do not fit in the representation of the initial cluster, recruit additional clusters, finally resulting in a set of competitive cluster each representing one category [6]. Another example of a hybrid model is RULEX, the rule-plus-exception-model [7], which assumes a stochastic process in which people can classify objects by forming simple rules with the addition of occasional exceptions. Importantly, different subjects form different simple rules and memorize different exceptions to those rules [7]. The main difference between the two aforementioned models is that RULEX can explain categorization based on two mutually exclusive categories, while SUSTAIN is intended to be a more general learning model [6].

These models try to explain category learning on a cognitive level [8] but make no strong predictions about neural substrates. More recent developments in cognitive neurosciences have tried to identify the neural basis of categorization processes, for example by using functional imaging or by performing comparative studies of humans and animals [2]. Various structures of the brain have been shown to participate in different forms of categorization learning, including visual association areas, the medial temporal lobe (MTL), the prefrontal cortex, and the basal ganglia, with the contribution of these structures depending on the experimental paradigm that is employed [1]. Furthermore, based on this widespread involvement of neural structures, it has been pointed out that it is improbable for categorization to be based on a single neural system, and that it requires the interaction of multiple brain structures and their plastic capabilities [9]. This view is also supported by convergent findings from neuroimaging studies that are not easily explainable by single-system approaches [10].

There has been a growing trend of integration in the two main categorization research fields, computational modeling and cognitive neuroscience, especially supported by the employment of neuroimaging studies. One example is the usage of computational models as the basis for the analysis of fMRI data [11].

Nevertheless, the neural basis of two of the most prominent category learning types, prototype- and exemplar-based learning, is yet unclear. The former might in part be mediated by the fusiform gyrus, as has been shown in a previous categorization study [12], where the activation of the fusiform gyrus changed after learning the membership to a category. Similarly, Pernet et al. [13] yielded evidence for fusiform gyrus activation in letter categorization. It has further been shown, that activation of the so called fusiform face area within the fusiform gyrus could reflect visual expertise [14,15], a process that is also involved in categorization learning.

On the other hand, exemplar-based learning could be processed by the MTL, since the explicit memorization of individual stimuli would require the involvement of memory systems that are tuned to sparse coding properties [16]. It has been shown previously that the activation of single neurons in the hippocampus can be linked to category-specific visual responses [16–18] and that cell firing within the hippocampus correlates with categorization performance [19]. Studies using formal modelling with SUSTAIN could also show an involvement of the MTL in a rule-plus-exception category learning task [11], emphasizing the role of the MTL in the mastering of exceptions to a category rule. The role of the hippocampus in stimulus generalization, representation and categorization has also been highlighted in a recent review [20], describing it as part of a network involving the basal ganglia and the prefrontal cortex and comparing its function to decision making processes.

A comparative study investigating the stages of category learning in humans and pigeons [21] successfully modeled prototype- and exemplar-based strategies over the course of an extended learning phase, showing that both species change their initial prototype-based strategies in order to correctly categorize stimuli that represent exceptions from the general similarity. With this, they also replicated previous studies investigating the time course of category learning [3,22,23]. Unfortunately, there was no investigation of the underlying neural basis for these strategies.

The present study aimed to investigate the neural correlates for prototype- and exemplar-based categorization strategies, in order to contribute to the question if both learning types are part of the same neural process or two individual and distinct processes, as well as to investigate the change of neural activation over the time course of the experiment. For this, we employed the same behavioral paradigm as Cook and Smith [21], in which participants had to categorize unfamiliar abstract stimuli into two groups by means of direct feedback after every trial. The categories consisted of a prototype, five typical stimuli and an exception (which shared more features with the opposing prototype).

Based on previous findings, we expected differential activation patterns for prototypes and typical stimuli on the one hand (prototype-based learning, mediated by the fusiform gyrus) and for exceptions on the other hand (exemplar-based learning, mediated by the hippocampus). Behaviorally, the learning of exceptions should be diminished in the beginning and should progressively increase over the course of the experiment, when participants realize that their prototype-based learning strategy does not work for the exceptions.

2. Material and methods

2.1. Participants

Twenty-eight right-handed and neurologically healthy subjects (12 male and 16 female subjects; mean age: 24.61 years; range: 20–30) participated in the experiment, reimbursed with research credit needed for their studies of psychology, or alternatively with 15€. All subjects gave informed written consent after a detailed explanation of the procedure. The study received ethical approval by the local Ethics Committee of the Medical Faculty of the Ruhr University Bochum, which conforms to the Declaration of Helsinki.

2.2. Stimuli and task

The experiment took place inside of an MRI scanner and was performed using Presentation® software and MRI video goggles with a resolution of 800 × 600 pixels, registering the responses with an MRI-suitable keypad. Participants had to perform a visual categorization task, which was adapted from Cook and Smith [21]. In this task, circular stimuli (400 × 400 pixels) with six binary color dimensions had to be categorized into one of two stimulus “families”, with the participants having no prior knowledge about the stimuli or categories. The stimuli were similar in their structure but differed in the color combinations. Each category consisted of one prototype, five typical stimuli that shared five colors with the prototype, and one exception that shared five colors with the prototype of the other category (see Fig. 1). This design prevented the usage of a prototype-based strategy for the exceptions, since this strategy would lead to an incorrect categorization.

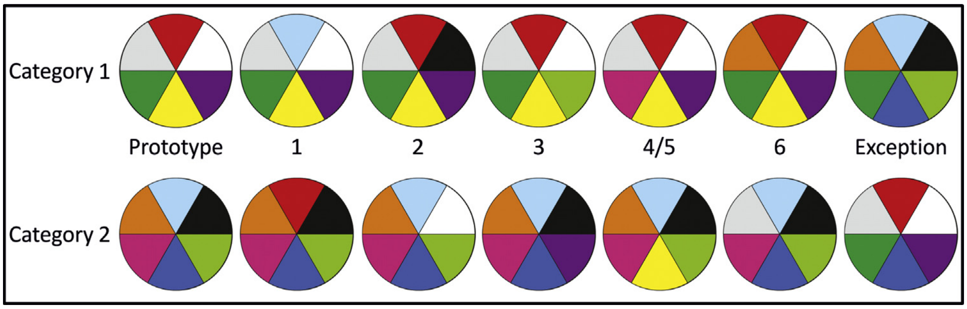


Fig. 1. Stimuli from both categories. All stimuli were constructed with 3.88 shared colors within and 2.12 shared colors between categories. Additionally, the stimuli of one category shared 4.57 colors with their own prototype and 1.43 with the prototype of the other category [21]. In general, one prototype alongside six typical stimuli was used in each category, with the fourth respective fifth typical stimulus being used as an exception for the other category.

All stimuli were presented centered on a white background. Responses were recorded using two buttons, with each button corresponding to one of the two categories. Immediately after the response feedback was given: "correct" or "incorrect". Alternatively, the feedback was "please react faster!" if the participant did not press a button within 2.2 s of stimulus presentation. The feedback was presented for 1 s, followed by a fixation cross that was presented for a variable period of 1–2 s before the next trial started. The experiment consisted of five blocks, with each block being composed of 98 trials, resulting in 490 trials over the course of the experiment. All stimuli were randomly presented in seven trials of each block; therefore each block contained 14 prototypes, 70 typical stimuli and 14 exceptions. Participants were allowed to make a pause after every block.

2.3. Image acquisition

The experiment was performed using a Philips 3 T Achieva MRI scanner with a 32- channel SENSE head coil. A T1 weighted structural scan was acquired for every participant at the start of the first experimental session (220 slices, voxel size = 1 × 1 × 1 mm, TE = 3.8 ms, flip angle = 8°). T2* weighted echo-planar MR images (EPI) were acquired in all three experimental conditions in an ascending sequence of 30 slices (voxel size = 1.65 × 1.65 × 5 mm, TR = 2200 ms, TE = 35 ms, flip angle = 90°, SENSE factor = 3). The first five images at the start of each session were discarded to allow for MRI signal stabilization.

2.4. Data analysis

For the analysis of the behavioral data, participants were divided into two groups: learners, who reached a correct response rate of at least 70% for all stimulus types in the last block, and non-learners with less than 70% correct responses for at least one stimulus type. After splitting the subjects into the two groups, additional *t*-tests were calculated to ensure that the behavioral performance of the non-learners did not differ significantly from the 50% chance level. While learners showed an increase of correct responses for all stimulus types over the five blocks (prototype: t = 13.1; p = 0.00, exception t = 16.5; p = 0.00), the non-learners remained at the chance level of about 50% (prototype: t = 1.35; p = 0.2, exception t = 0.3; p = 0.7) during the fifth block.

Data from prototypes and typical stimuli were combined and were both treated as prototypes, since each of the typical stimuli could structurally represent the prototype for the other stimuli. A repeated measures ANOVA (with Greenhouse Geisser correction) with the factors "stimulus" (ProTyp vs. Exc.), "block" (1; 2; 3; 4; 5), and the inner-subject factor "learner" (yes vs. no) was applied to the correct responses.

The imaging data was preprocessed using the latest release of SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8). The preprocessing consisted of slice-time correction, realignment (with unwarping), co-registration of the EPIs with the structural scan, segmentation of the structural scans into grey and white matter, and normalization to MNI space using DARTEL [24]. EPIs were resliced into 2 × 2 x 2 mm voxels, and finally smoothed with a Gaussian kernel of 8 mm full-width half-maximum (FWHM). The preprocessed images were then submitted into a first level GLM analysis, where the blood oxygenation level dependency (BOLD) signal was modeled with the canonical hemodynamic response function. A high-pass filter at 128 s was used to remove low frequency drifts. All reported statistics refer to whole brain analyses. The statistical maps were thresholded at $p < 0.05$ using false discovery rate (FDR) correction for multiple comparisons [25], with a minimum of 15 contiguous voxels per cluster. Anatomical labeling was performed using the automated anatomical labeling toolbox (AAL [26]).

For the first level analysis, seven regressors were defined per block (resulting in 35 regressors), representing correct and incorrect responses for the three stimulus types ("correct prototypes", "incorrect prototypes", "correct typical", "incorrect typical", "correct exception", "incorrect exception") as well as an additional regressor for the fixation period. For the contrasts of interest, prototypical and typical stimuli were combined ("ProTyp"). Correct responses were contrasted against the fixation regressor, separate for block 1 (before learning) and block 5 (after learning), resulting in the contrasts "ProTyp untrained", "ProTyp trained", "Exc untrained", and "Exc trained". These four contrasts were then used for group inference in the second level analysis.

In the group-level random effects analysis, a within-subjects full factorial model with 3 factors was employed. The model included the factors "group" (learners vs. non-learners), "stimulus" (ProTyp vs. Exc.), and "block" (1 vs. 5). Main effects for all factors and interactions were used as contrasts of interest. Selected significant activation foci from interaction contrasts were used to extract mean signal changes (in percent) with MarsBaR software (http://marsbar.sourceforge.net/), and were then fed into paired *t*-tests.

To further examine the time-course of activation in different brain structures, Pearson correlations were computed for the mean signal changes of the ROIs that were defined from results of the random effects analysis as described above (in this case the ROI covered activation in the hippocampus and fusiform gyrus). Correlations of the mean percent signal changes that have been extracted from the hippocampus and the fusiform gyrus, as described above by using the MarsBaR software package, were computed for learners and non-learners and for both stimulus types separately for the course of five blocks.

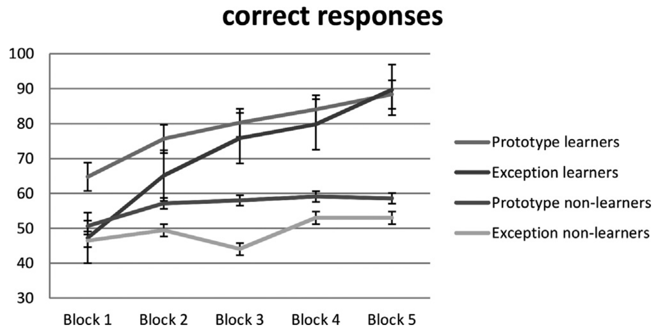


Fig. 2. Correct responses in the categorization task. Learners show an increase in correct responses over the five experimental blocks for both stimulus types, while non-learners do not show any change. Error bars represent SEM.

3. Results

3.1. Behavioral data

The post hoc differentiation of the participants into two groups yielded 17 learners and 11 non-learners. The repeated measures ANOVA revealed significant main effects for the factors "stimulus" ($F_{(1,26)} = 4.502$; $p < 0.05$) and "block" ($F_{(4,104)} = 15.138$; $p < 0.0001$), and a significant interaction between the factors "block" and "learner" ($F_{(4,104)} = 8.055$; $p < 0.0001$). Paired *t*-tests showed that the interaction resulted from the different learning curves of both groups (Fig. 2). Learners and non-learners did not differ at the first block ($t = -1.8$; $p = 0.07$) but they differed on the second till the fifth block (Block 2: $t = -3.2$; $p < 0.001$, Block 3: $t = -4.7$; $p < 0.001$, Block 4: $t = -3.6$; $p < 0.001$, Block 5: $t = -6.9$; $p < 0.001$).

3.2. Imaging data

The full factorial analysis yielded significant activation clusters for the main effects "group" and "block".

For the main effect of "group", three activation clusters were found for the contrast learners > non-learners (see Fig. 3): one in the right (14 −50 0, 208 voxels, T = 6.18) and the left lingual gyrus (−16 −60 4, 66 voxels, T = 4.82) as well as one in the left inferior frontal gyrus, extending to the insula and temporal pole (−46 16 −4, 53 voxels, T = 4.85). The main effect of the factor "block" yielded activations spanning the entire brain and will not be considered here in detail (see Table 1 for full details). However, there was a large significant activation cluster in the left striatum (−10 4 18, 667 voxels, T = 4.66), with the peak value in the caudate nucleus.

Additionally, the interaction contrast of "block" x "group" yielded significant activations (see Fig. 4 and Table 1), with clusters in the left (−40 12 −16, 64 voxels, T = 4.73) and right temporal pole (34 10 −26, 100 voxels, T = 4.70), the pars triangularis of the left inferior frontal gyrus (−54 20 −2, 44 voxels, T = 4.65), and the left fusiform gyrus (−42 −48 −22, 24 voxels, T = 4.23).

For further analysis, the interaction was divided into separate contrasts for each of the stimulus types, resulting in an additional significant cluster in the right hippocampus (30 −24 −10, 23 voxels, T = 3.97) and left parietal lobe (−50 −60 40, 76 voxels, T = 3.94) for exceptions (see Table 1 for a detailed listing). The interaction contrast for prototype stimuli did not show any significant activation. Signal changes were extracted for the two areas of interest, the left fusiform gyrus and the right hippocampus. T-tests revealed that the interactions were based on a higher activation for learners in the last block compared to the first block, with a significantly higher signal change difference for prototype stimuli in the fusiform gyrus ($T_{(1,16)} = 2.435$; $p < 0.05$) and a trend for the signal change difference of exceptions in the right hippocampus ($T_{(1,16)} = 2.098$; $p = 0.062$; see Fig. 5). The percent signal change of non-learners did not show any significant differences between the blocks or stimulus types (Fig. 5).

The time course over the five blocks of each correlation between the mean signal change in the hippocampus and the fusiform gyrus is presented in Fig. 6. This time course shows the interplay between these two regions. The analysis of correlations of the mean signal change in the right hippocampus and left fusiform gyrus revealed that successful learners showed an early correlation of signal change between both structures for the learning of prototypes. For exceptions, the correlation only became significant in the last part of the experiment (see Fig. 6). Notably, there was a drop of correlation in the second block of the experiment, resulting in approximately zero correlation between both structures and for both stimulus types. This drop occurred due to an earlier positivity of the hippocampal signal change for both stimulus types in comparison with the left fusiform gyrus. In general, these correlations showed a time displaced course for learning prototypes and exceptions with a later involvement of the fusiform gyrus in the successful learning of the exceptions. Additionally, there was almost no significant correlation for the non-learners, with only a small correlation ($p < 0.05$) for exceptions in the last block of the experiment.

4. Discussion

The aim of this study was to reveal the neural correlates of prototype- and exemplar-based category learning and to investigate how activations in the involved brain regions change over an extended learning period. For this, participants had to perform a categorization task similar to the paradigm employed by Cook and Smith [21]. After showing an unfamiliar circular stimulus with six binary color dimensions, participants had to decide without any prior knowledge to which one of two categories the stimulus belonged. Direct feedback was given after every trial to enable learning of the category structure. In order to disrupt the abstraction-based strategy used for the prototype (prototypes and typical combined) stimuli, each of the two categories contained one exception. In order to successfully categorize these exceptions, participants had to explicitly memorize these items and hence had to use an exemplar-based strategy.

The behavioral data in the current study replicated previous results [21]. Overall, the performance was slightly poorer, although it has to be noted that the original study only tested and reported behavioral results of nine human participants. Here, 17 out of 28 subjects were classified as learners, based on a correct response count of higher than 70% in the last experimental block. A possible explanation for those subjects below this criterion might be the unusual environment and the noise caused by the MRI scanner. Learners showed a good performance for prototype stimuli early on, with a steady increase over the five experimental blocks. Correct categorizations of exceptions started at chance level in the first block but showed a rapid increase after the second block, with a comparable performance to prototype stimuli in the last block. The occurrence of subjects with different learning performances presented the opportunity to study the processes that differentiate between levels of category learning.

The analysis of the imaging data revealed significant activations of the lingual gyrus and the left inferior frontal gyrus when comparing learners and non-learners, regardless of block or stimulus type. Thus, learners possibly performed at a higher level since they successfully activated the lingual gyrus, an area that is involved in analyzing and memorizing visual color stimuli [27]. The activation of Broca's area is very likely related to inner verbalizations [28] in a possible attempt to acquire knowledge about category mem-

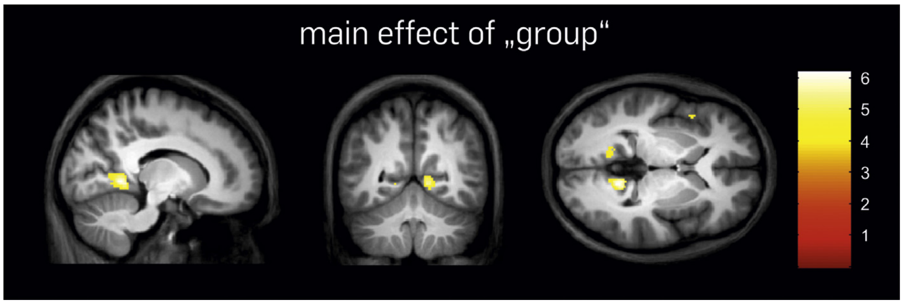


Fig. 3. Significant activations for the main effect of “group” in the full factorial design. Note that the main effect of “stimulus” did not yield any significant activation. All activations are FDR corrected for multiple comparisons, with p < 0.05.

Table 1
Significant activations (p < 0.05, FDR-corrected) for main effects and interactions in the full factorial analysis. Each line represents one cluster, with anatomical labels from AAL.

Contrast	MNI coordinates (x, y, z)			Cluster size	T	Anatomical structure
Block: Block 5 > Block 1	−20	−96	−6	422	5.76	l. Inferior occipital gyrus
	−40	12	−16	2329	5.48	l. Inferior frontal gyrus/insula/temporal pole
	−44	−50	−20	8088	5.29	l. & r. temporal lobe, peak in fusiform gyrus
	−10	4	18	667	4.66	l. caudate nucleus
	2	−24	68	794	4.41	medial frontal gyrus/precentral/postcentral gyrus
	20	−92	−6	260	4.15	r. inferior occipital gyrus
	2	24	56	166	3.88	supplementary motor area
	0	46	32	388	3.77	medial frontal gyrus
	−52	28	16	34	3.56	l. inferior frontal gyrus, pars triangularis
	44	−48	−18	110	3.51	r. fusiform gyrus
	−58	−12	30	82	3.50	l. precentral gyrus
	2	34	50	41	3.46	superior frontal gyrus
	−2	40	−4	104	3.34	l. anterior cingulate cortex
	50	−16	12	26	3.14	r. rolandic operculum
Group: Learners > Non-Learners	14	−50	0	208	6.18	r. lingual gyrus
	−46	13	−4	53	4.85	l. Inferior frontal gyrus/insula/temporal pole
	−16	−60	4	66	4.82	l. lingual gyrus
Interaction “block” x “group”	−40	12	16	64	4.73	l. temporal pole
	34	10	−26	100	4.70	r. temporal pole
	−54	20	−2	44	4.65	l. inferior frontal gyrus, pars triangularis
	−42	−48	−22	24	4.23	l. fusiform gyrus
Interaction “block” x “group-exceptions only	−54	22	−6	138	4.86	l. inferior frontal gyrus, pars triangularis
	−40	14	−14	100	4.76	l. inferior frontal gyrus
	42	14	−24	167	4.41	superior temporal gyrus/temporal pole
	54	−60	20	90	4.09	r. superior/middle temporal gyrus
	−42	−66	28	72	4.07	l. angular gyrus
	54	−60	20	90	4.09	r. superior/middle temporal gyrus
	30	−24	−10	23	3.97	r. hippocampus
	−50	−60	40	76	3.94	l. inferior parietal gyrus
	−2	−80	42	41	3.75	l. precuneus

bership. With these activations representing the only difference between learners and non-learners, one might assume that non-learners failed to utilize these processes and hence did not learn to correctly categorize the stimuli.

As hypothesized, activations of the hippocampus and the fusiform gyrus were found for exemplar- and prototype-based strategies, respectively, with these activations being apparent in the interaction contrasts of “block” and “group”. The mean signal change in these clusters was extracted to provide insight into the neural processes involved in the categorization of different stimulus types before and after learning the category structure. Distinct neural correlates for the two categorization strategies emerged, with the left fusiform gyrus being more involved in prototype-based learning, and the right hippocampus being activated for the exemplar-based learning of exceptions.

These results are in line with previous findings. Gauthier et al. [14] showed that an extensive training with novel objects (“greebles”) leads to activation of a “greeble”-specific area in the fusiform gyrus. Also, expertise for other objects can lead to activations in the fusiform gyrus and occipital lobe [29]. This is possibly due to local changes of synaptic strengths, lowering thresholds for recognizing target stimulus elements and thereby increasing classification of a coherent object “family”, regardless of individual stimuli [30]. The approaches by Gauthier et al. [29] and the present one differ with respect to the time that was available for the subjects to acquire expertise. The stimuli of the studies (greebles and circles) did also differ with respect to the features and feature organization. The explanation given by Gauthier and co-worker might therefore provide a fist explanation for the current findings which has to be seen with its limitations. However, these changes likely involve a population coding account of object configurations [31]. Even further proof for the importance of the left visual ventral stream was found in an examination of patients with lesions to the left posterior hemisphere. They did show more deficits in generalization and visual category learning than patients with right posterior cerebral lesions [32] showing the importance of visual association areas for category learning processes.

While the contribution of the fusiform gyrus possibly enabled the participants to correctly categorize prototypes and similar stimuli, the results differed for the learning of exceptions. After a few

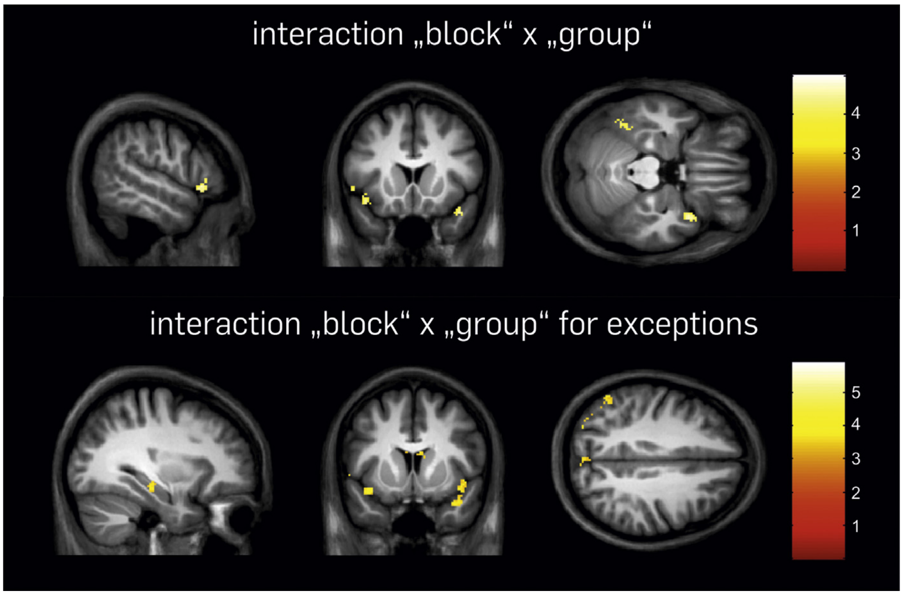

interaction „block“ x „group“
4
3
2
1
interaction „block“ x „group“ for exceptions
5
4
3
2
1

Fig. 4. Significant activations for two interactions. Note that the interaction “block” x “group” for prototype stimuli did not yield significant activations with the FDR-corrected threshold of p < 0.05.

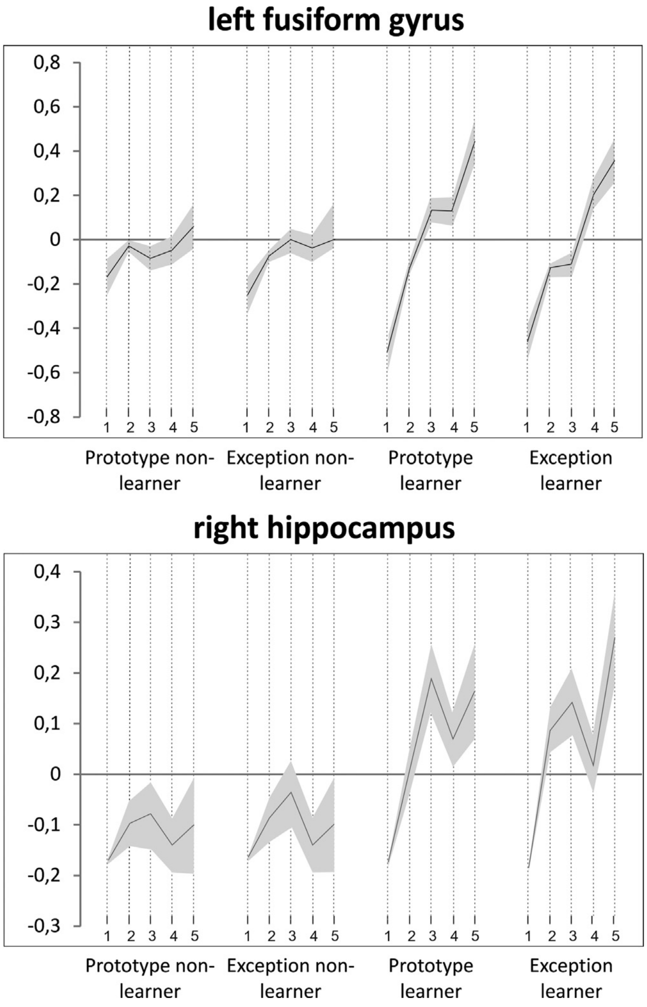


Fig. 5. Percent signal change of two significant activation clusters. For both ROIs, *t*-tests revealed significant differences for learners between the first and last learning session, as well as higher signal change differences for prototype stimuli in the left fusiform gyrus and a trend for exceptions in the right hippocampus.

blocks, the participants realized that their prototype-based strategy used for 12 of 14 stimuli did not work for the two exceptions and switched to an exemplar-based strategy, as previously shown by Cook & Smith [21]. This strategy required an explicit categorization and memorization of the exceptions, which was possibly enabled by the category-sensitive cells in the MTL, especially the hippocampus [17,19]. The hippocampal activation shown in the interaction contrast and also for the main effect of “block” (see Table 1), did distinguish between successful learners of the exceptions in the last block.

Indeed, MTL neurons in monkeys and humans show selective responses to classes of visual stimuli and to specific individuals [33]. These results suggest an invariant, sparse and explicit code [16]. Coding properties of hippocampal neurons are highly flexible and quickly learn to categorize visual stimuli by extracting unique combination of features that are relevant for discrimination [19]. These experiments reveal that hippocampal neurons are able to code for individual items using a multitude of different representations. These neurons are sparse in the sense that they fire after the presentation of only very few stimuli [34]. It is probable that the sparse coding is supported by the firing of inhibitory neurons, resulting in an increase of the hemodynamic response and hence the BOLD signal [35]. Inhibition could modify synaptic connections in the MTL via long-term depression and lead to more specific neural representations [36].

One argument against a vital contribution of the hippocampus to categorization learning are the findings that show that categorization learning can be intact although the hippocampus is damaged and recognition memory is impaired in patients [37–39]. However, this dissociation may also be based on differential memory demands and possibly on residual resources [40–42].

The inferior frontal gyrus frontal gyrus showed also increased activation when comparing the first and the fifth block. This is in line with different findings in the literature which emphasize the critical contribution to categorization. Wallis and Miller have shown nearly 10 years ago in monkeys, that the premotor cortex is involved in the retrieval and application of abstract rules [43], whereas Halsband and Passingham could show that damage to the premotor cortex disrupts the response to previously learned stimuli [44]. These results emphasize the role of the premotor cortex in categorization and give a good explanation for the inferior frontal

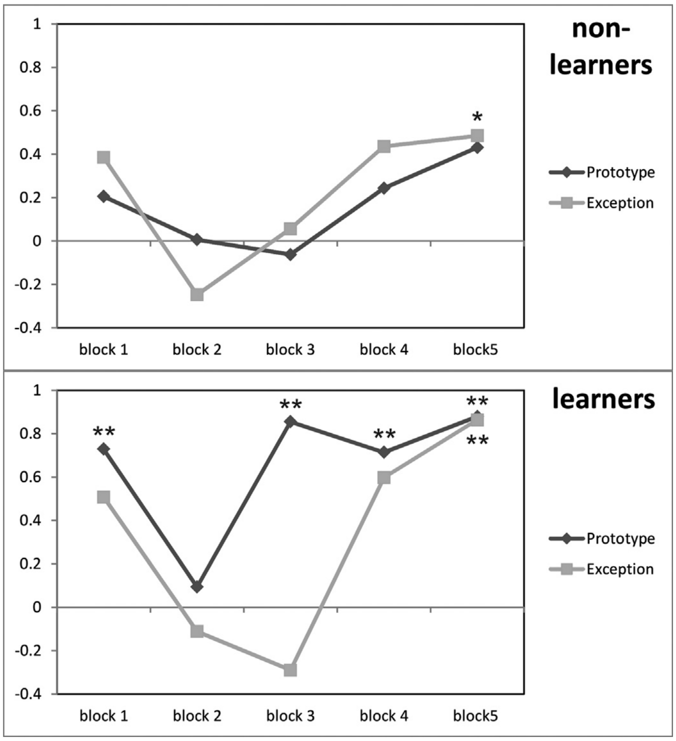


Fig. 6. Correlations of percent signal change in hippocampus and fusiform gyrus. Pearson correlations were calculated separately for learners and non-learners, with one asterisk representing a significance threshold of p < 0.05 and two asterikses representing p < 0.01.

gyrus activation in the present experiment. Present results support earlier findings in categorization learning and extend them into the context of categorization learning of prototype and exception.

Seger and Peterson [20] describe the basal ganglia and their connections to the hippocampus as well as parieto-frontal networks as being essential for categorization, especially since these structures all receive dopaminergic projections which are essential for the coding of reward [20]. Additionally, the basal ganglia play an essential role in sensory integration and response selection or thresholding [20,45], with the input being received through striatal nuclei. In line with this, the main effect of “block” did show a significant activation of the caudate nucleus, showing the importance of this structure for the successful learning over time, which is also true for learning beyond categorization, e.g. habit learning and automaticity [46]. Findings from research in monkeys suggested that the contribution of the basal ganglia to categorization precedes cortico-cortical activation [47] which is not supported by the present findings demonstrating caudate activation at a late stage of learning. This might be explained by the cortico-striatal loop involved in visual associative processing as suggested by Seger [48].

Up to now, the current data could be misunderstood as pointing to a complete dichotomy in the processing of prototypes and exemplars. However, the correlation of the mean signal changes over time revealed a synchronization of activation changes between right hippocampus and left fusiform gyrus. The activations were already synchronized in the first block, with the learners showing negative signal changes for both stimulus types. The desynchronization after the first block revealed a different progression for the signal changes of both structures, with the hippocampus showing more positive signal changes than the fusiform gyrus in the second block. In the third block, the correlation of both structures became significant again for prototypes, followed by exceptions in the last block. In general, the hippocampal activation showed a parallel, but time displaced activation course for exception and prototype learners. The activation of the fusiform gyrus, which might reflect the representation of the stimulus, emerged later for the exceptions than the prototype, which is also reflected by the behavioral data. Taken together, the population coding properties of the visual association cortex in the fusiform gyrus possibly enabled the learners to build representations of the stimulus families with shared similarities among their features. The increase of activation over time corresponded to the ability to successfully categorize prototype stimuli early on during the experiment. It also enabled participants to discriminate the exceptions from the rule, in interplay with the sparse coding properties of the hippocampus, which coded the representation of exceptions. These processes were additionally supported by frontal and striatal activations that enabled reward prediction, action selection, and thresholding. Further evidence for the hippocampus not being exclusively involved in exemplar learning but also in the learning of prototypes, comes from an fMRI study that differentiated between A/B and A/non-A prototype learning [49]. In this study, the authors could demonstrate a hippocampal contribution specifically for A/B prototype learning tasks, which is also in agreement with the current results.

It is important to note that the current study was conceived and conducted within a framework of theories that posit prototype and exemplar processes as different constituents of category learning [1,3,50–53], in order to integrate computational models with modern cognitive neuroscience methods. However, recent studies have also demonstrated that the underlying cognitive processes cannot be neatly separated [54]. In line with this, common element models (see Soto & Wassermann [55] for a review and mathematical formulation) assume that most stimuli have some common elements that can make them similar. Thus, the perceptual similarity between two stimuli is a direct function of the proportion of shared elements. Conversely, non-shared features drive dissimilarity. The common elements model can be extended to the current experiment as well as the study of Cook & Smith [21] by assuming that our exception exemplar represents nothing else than a unique cue within the common elements framework. Accordingly, each stimulus feature employed in a categorization task is associated independently with an outcome and activates an individual configural unit which represents that unique combination [56]. Such an account would predict faster learning of stimuli with more common elements due to shared associative strength and slower learning of exception stimuli [55]. Nevertheless, a recent study using multivariate pattern analysis [57] could demonstrate that exemplar-based models cannot be discarded as of yet, since the MVPA did reveal activation patterns that are best explained by exemplar-based models.

5. Conclusion

Taken together, results from the current study reveal two neural substrates that are associated with a prototypical stimulus that shares a large number of common elements with other members of a class and an exceptional exemplar that represents a unique cue. While the former seems to require object-based processes of the ventral visual stream, the latter additionally needs the hippocampus to create a sparse code for a stand-alone representation of the exceptions. However, while the current data are in line with the assumptions of a multiple system model that incorporates differing prototype- and exemplar based learning strategies, there is no complete dichotomy for the involvement of both structures. In order to further elucidate the underlying systems, future research might focus on formal modeling of the initial assumptions and the validation of these models by using MVPA to disentangle the patterns of activation that enable successful category learning.

Conflict of interest

The author(s) declare no competing financial interests.

Acknowledgments

The present study was funded by a grant (Sonderforschungsbereich 874, CRC 874) from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) to Boris Suchan (Project B8) and Onur Güntürkün (Project B5). Robert Lech is supported by the International Graduate School of Neuroscience and the Ruhr-University Research School. We thank Sabine Bierstedt for the preparation of the stimulus material and for her help during data acquisition.

References

[1] F.G. Ashby, W.T. Maddox, Human category learning, Annu. Rev. Psychol. 56 (2005) 149–178.
[2] J.D. Smith, M.E. Berg, R.G. Cook, M.S. Murphy, M.J. Crossley, J. Boomer, et al., Implicit and explicit categorization: a tale of four species, Neurosci. Biobehav. Rev. 36 (2012) 2355–2369.
[3] J.D. Smith, J.P. Minda, Prototypes in the mist: the early epochs of category learning, J. Exp. Psychol.: Learn. Mem. Cogn. 24 (1998) 1411.
[4] J.K. Kruschke, ALCOVE: an exemplar-based connectionist model of category learning, Psychol. Rev. 99 (1992) 22–44.
[5] R.M. Nosofsky, T.J. Palmeri, An exemplar-based random walk model of speeded classification, Psychol. Rev. 104 (1997) 266–300.
[6] B.C. Love, D.L. Medin, T.M. Gureckis, SUSTAIN: a network model of category learning, Psychol. Rev. 111 (2004) 309–332.
[7] R.M. Nosofsky, T.J. Palmeri, S.C. McKinley, Rule-plus-exception model of classification learning, Psychol. Rev. 101 (1994) 53–79.
[8] S. Lewandowsky, T.J. Palmeri, M.R. Waldmann, Introduction to the special section on theory and data in categorization: integrating computational, behavioral, and cognitive neuroscience approaches, J. Exp. Psychol. Learn. Mem. Cogn. 38 (2012) 803–806.
[9] C.A. Seger, E.K. Miller, Category learning in the brain, Annu. Rev. Neurosci. 33 (2010) 203–219.
[10] R.A. Poldrack, K. Foerde, Category learning and the memory systems debate, Neurosci. Biobehav. Rev. 32 (2008) 197–205.
[11] T. Davis, B.C. Love, A.R. Preston, Learning the exception to the rule: model-based FMRI reveals specialized representations for surprising category members, Cereb. Cortex 22 (2012) 260–273.
[12] J.R. Folstein, T.J. Palmeri, I. Gauthier, Category learning increases discriminability of relevant object dimensions in visual cortex, Cereb. Cortex 23 (2013) 814–823.
[13] C. Pernet, P. Celsis, J. Demonet, Selective response to letter categorization within the left fusiform gyrus, Neuroimage 28 (2005) 738–744.
[14] I. Gauthier, M.J. Tarr, A.W. Anderson, P. Skudlarski, J.C. Gore, Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects, Nat. Neurosci. 2 (1999) 568–573.
[15] M.H. Tong, C.A. Joyce, G.W. Cottrell, Why is the fusiform face area recruited for novel categories of expertise? A neurocomputational investigation, Brain Res. 1202 (2008) 14–24.
[16] R.Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, Invariant visual representation by single neurons in the human brain, Nature 435 (2005) 1102–1107.
[17] G. Kreiman, C. Koch, I. Fried, Category-specific visual responses of single neurons in the human medial temporal lobe, Nat. Neurosci. 3 (2000) 946–953.
[18] B.A. Olshausen, D.J. Field, Sparse coding of sensory inputs, Curr. Opin. Neurobiol. 14 (2004) 481–487.
[19] R.E. Hampson, T.P. Pons, T.R. Stanford, S.A. Deadwyler, Categorization in the monkey hippocampus: a possible mechanism for encoding information into memory, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 3184–3189.
[20] C.A. Seger, E.J. Peterson, Categorization = decision making + generalization, Neurosci. Biobehav. Rev. 37 (2013) 1187–1200.
[21] R.G. Cook, J.D. Smith, Stages of abstraction and exemplar memorization in pigeon category learning, Psychol. Sci. 17 (2006) 1059–1067.
[22] M. Blair, D. Homa, Expanding the search for a linear separability constraint on category learning, Mem. Cogn. 29 (2001) 1153–1164.
[23] J.D. Smith, W.P. Chapman, J.S. Redford, Stages of category learning in monkeys (Macaca mulatta) and humans (Homo sapiens), J. Exp. Psychol. Anim. Behav. Process. 36 (2010) 39–53.
[24] J. Ashburner, A fast diffeomorphic image registration algorithm, Neuroimage 38 (2007) 95–113.
[25] C.R. Genovese, N.A. Lazar, T. Nichols, Thresholding of statistical maps in functional neuroimaging using the false discovery rate, Neuroimage 15 (2002) 870–878.
[26] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, et al., Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, Neuroimage 15 (2002) 273–289.
[27] X. Wang, Z. Han, Y. He, A. Caramazza, L. Song, Y. Bi, Where color rests: spontaneous brain activity of bilateral fusiform and lingual regions predicts object color knowledge performance, Neuroimage 76 (2013) 252–263.
[28] S.S. Shergill, E.T. Bullmore, M.J. Brammer, S.C. Williams, R.M. Murray, P.K. McGuire, A functional study of auditory verbal imagery, Psychol. Med. 31 (2001) 241–253.
[29] I. Gauthier, P. Skudlarski, J.C. Gore, A.W. Anderson, Expertise for cars and birds recruits brain areas involved in face recognition, Nat. Neurosci. 3 (2000) 191–197.
[30] F.A. Soto, E.A. Wasserman, A category-overshadowing effect in pigeons: support for the Common Elements Model of object categorization learning, J. Exp. Psychol. Anim. Behav. Process. 38 (2012) 322–328.
[31] T. Hirabayashi, Y. Miyashita, Dynamically modulated spike correlation in monkey inferior temporal cortex depending on the feature configuration within a whole object, J. Neurosci. 25 (2005) 10299–10307.
[32] B. Langguth, M. Jüttner, T. Landis, M. Regard, I. Rentschler, Differential impact of posterior lesions in the left and right hemisphere on visual category learning and generalization to contrast reversal, Neuropsychologia 47 (2009) 2927–2936.
[33] F. Mormann, S. Kornblith, R.Q. Quiroga, A. Kraskov, M. Cerf, I. Fried, et al., Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe, J. Neurosci. 28 (2008) 8865–8872.
[34] R.Q. Quiroga, G. Kreiman, C. Koch, I. Fried, Sparse but not 'grandmother-cell' coding in the medial temporal lobe, Trends Cogn. Sci. (Regul. Ed.) 12 (2008) 87–91.
[35] I.V. Viskontas, B.J. Knowlton, P.N. Steinmetz, I. Fried, Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions, J. Cogn. Neurosci. 18 (2006) 1654–1662.
[36] N. Axmacher, C.E. Elger, J. Fell, Memory formation by refinement of neural representations: the inhibition hypothesis, Behav. Brain Res. 189 (2008) 1–8.
[37] B.J. Knowlton, L.R. Squire, The learning of categories: parallel brain systems for item memory and category knowledge, Science 262 (1993) 1747–1749.
[38] L.R. Squire, B.J. Knowlton, Learning about categories in the absence of memory, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 12470–12474.
[39] A. Bozoki, M. Grossman, E.E. Smith, Can patients with Alzheimer's disease learn a category implicitly? Neuropsychologia 44 (2006) 816–827.
[40] R.M. Nosofsky, S.E. Denton, S.R. Zaki, A.F. Murphy-Knudsen, F.W. Unverzagt, Studies of implicit prototype extraction in patients with mild cognitive impairment and early Alzheimer's disease, J. Exp. Psychol. Learn. Mem. Cogn. 38 (2012) 860–880.
[41] R.M. Nosofsky, S.R. Zaki, Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation, Psychol. Sci. 9 (1998) 247–255.
[42] S.R. Zaki, R.M. Nosofsky, A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli, Cogn. Affect. Behav. Neurosci. 1 (2001) 344–359.
[43] J.D. Wallis, E.K. Miller, From rule to response: neuronal processes in the premotor and prefrontal cortex, J. Neurophysiol. 90 (2003) 1790–1806.
[44] U. Halsband, R. Passingham, The role of premotor and parietal cortex in the direction of action, Brain Res. 240 (1982) 368–372.
[45] F.A. Soto, E.A. Wasserman, Mechanisms of object recognition: what we have learned from pigeons, Front. Neural Circuits (2014) 8.
[46] F.G. Ashby, B.O. Turner, J.C. Horvitz, Cortical and basal ganglia contributions to habit learning and automaticity, Trends Cogn. Sci. 14 (2010) 208–215.
[47] F.G. Ashby, J.M. Ennis, B.J. Spiering, A neurobiological theory of automaticity in perceptual categorization, Psychol. Rev. 114 (2007) 632–656.
[48] C.A. Seger, How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback, Neurosci. Biobehav. Rev. 32 (2008) 265–278.
[49] D. Zeithamova, W.T. Maddox, D.M. Schnyer, Dissociable prototype learning systems: evidence from brain imaging and behavior, J. Neurosci. 28 (2008) 13194–13201.
[50] D.L. Medin, M.M. Schaffer, Context theory of classification learning, Psychol. Rev. 85 (1978) 207–238.
[51] M.I. Posner, S.W. Keele, On the genesis of abstract ideas, J. Exp. Psychol. 77 (1968) 353–363.
[52] J.D. Smith, J.P. Minda, Journey to the center of the category: the dissociation in amnesia between categorization and recognition, J. Exp. Psychol. Learn. Mem. Cogn. 27 (2001) 984–1002.
[53] D.L. Medin, P.J. Schwanenflugel, Linear separability in classification learning, J. Exp. Psychol.: Hum. Learn. Mem. 7 (1981) 355–368.
[54] T. Davis, B.C. Love, A.R. Preston, Striatal and hippocampal entropy and recognition signals in category learning: simultaneous processes revealed by model-based fMRI, J. Exp. Psychol. Learn. Mem. Cogn. 38 (2012) 821–839.
[55] F.A. Soto, E.A. Wasserman, Error-driven learning in visual categorization and object recognition: a common-elements model, Psychol. Rev. 117 (2010) 349–381.
[56] M.A. Gluck, Stimulus generalization and representation in adaptive network models of category learning, Psychol. Sci. 2 (1991) 50–55.
[57] M.L. Mack, A.R. Preston, B.C. Love, Decoding the brain's algorithm for categorization from its neural implementation, Curr. Biol. 23 (2013) 2023–2027.