Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

NeuroImage 52 (2010) 699–709

Contents lists available at ScienceDirect

NeuroImage

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

Differential category learning processes: The neural basis of comparison-based learning and induction

Rubi Hammer [a,b,e,*], André Brechmann [d], Frank Ohl [d], Daphna Weinshall [a,c], Shaul Hochstein [a,b]

[a] Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel
[b] Neurobiology Department, Institute of Life Sciences, Hebrew University, Jerusalem, Israel
[c] School of Computer Sciences and Engineering, Hebrew University, Jerusalem, Israel
[d] Leibniz Institute for Neurobiology, Magdeburg, Germany
[e] Department of Psychology, Stanford University, Stanford, CA, USA

ARTICLE INFO

Article history:
Received 14 December 2009
Revised 22 February 2010
Accepted 29 March 2010
Available online 2 April 2010

Keywords:
Category learning
Categorization
Rule-based
Induction
Basal ganglia
Dorsal striatum
Hippocampus
Visual cortex

ABSTRACT

Findings from numerous studies suggest that multiple neural systems are involved in category learning. Specifically, it is often argued that acquiring a representation of different category structures (e.g., rule-based vs. prototype-based representation) involves different computational challenges, which are resolved by different neural circuitries in the human brain. Here we present an alternative approach for studying neural mechanisms of category learning: We refer to the idea that any category learning task involves mapping common features shared by same-category members, distinctive features discriminating members of different categories, or both. We argue that since these processes are psychologically and computationally distinct, they differ in their usability for category learning. Our participants learned novel categories of complex visual stimuli by comparing either pairs of objects from the same novel category or pairs of objects from different categories. Object pairs were chosen so that the objective amount of information they contained was identical in the two category learning conditions, equally enabling learning the predefined objective category structure. We find that the neural circuitry involved in detecting important between-categories differences is associated mainly with the dorsal striatum (bilaterally) and the right hippocampus. On the other hand, mapping within-category similarities and differences is restricted to high-level visual brain areas. We suggest that multiple neural mechanisms are involved in category learning enabling us to face different computational challenges associated with different basic types of induction processes that differ in their usability for learning different category structures.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Category learning is the cognitive ability that enables creation of compact and meaningful mental representations of objects and events. Evidence from behavioral studies suggests that category learning involves more than one mental process: In some scenarios, a categorization rule can be described verbally as a set of feature-dimensions by which object categories can be formed or discriminated (e.g., blue balls vs. red cubes). On the other hand, learning scenarios such as information-integration category learning, (where there is a within-category interdependence among certain feature-dimensions), or some prototype learning tasks, involve learning representations which people fail to describe verbally (Ashby and Gott, 1988). For this reason it is frequently suggested that learning different category structures involves a variety of computational challenges, which are met by different neural mechanisms in the human brain (Ashby and O'Brien, 2005; Smith and Grossman, 2008).

We now present an alternative approach for studying the nature of the neural mechanisms involved in category learning, while addressing previous findings showing that multiple neural mechanisms are involved in category learning. Our findings suggest that different neural mechanisms may be engaged for the task of learning the very same objective category structure, depending on the nature of the learning context and the provided information. Specifically, we find that when healthy adult participants learn a novel categorization rule (when categorizing novel complex visual stimuli), the processing of informative *between-category* differences, (as when comparing objects from *different* categories), is associated with enhanced neural activity in the dorsal striatum (bilaterally), the right hippocampus, and left lingual gyrus (ventral BA-19). In contrast, when the same categorization rule is learned by processing similarly informative *within-category* similarities and differences, (as when comparing objects from the same category), the neural activity in these brain areas is for the most part no higher than in a task that does not involve learning. Instead, we provide evidence showing that the process of

* Corresponding author. Department of Psychology, Jordan Hall (building 420), Stanford University, Stanford, CA 94305, USA.
E-mail address: rubih@stanford.edu (R. Hammer).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.03.080

Author's personal copy

mapping *within-category* similarities and differences is restricted to high-level visual brain areas including the right inferior temporal (BA-37) and the right angular (BA-39) gyri.

Early evidence for the existence of multiple neural mechanisms for category learning came from studying people with neural pathologies. Impaired neostriatum function in patients with Parkinson's Disease (PD) was reported to be associated with poor procedural learning, whereas damage to the limbic-diencephalic region in amnesic patients was reported to be associated with poor declarative learning (Knowlton et al., 1996). This account is supported by more recent neuroimaging studies with healthy participants (e.g., Nomura et al., 2007). Moreover, the striatum seems to play multiple roles in category learning: Some studies show that PD patients suffer mainly from poor abilities in information-integration category learning tasks which involve implicit learning (Maddox and Filoteo, 2001), whereas others suggest that PD patients suffer mainly from poor declarative rule learning (Ashby et al., 2003). More recently, by testing both PD patients' behavior, and by testing healthy participants using neuroimaging techniques, Filoteo and colleagues (2005a,b) showed that in the context of supervised learning, the striatum may be involved in directing attention to the task-relevant feature-dimensions by filtering out those that are task-irrelevant. According to Cincotta and Seger (2007), subparts of the striatum play different roles in category learning so that the head of the caudate is associated with feedback processing, while the body and tail of the caudate and the putamen are involved in creating stimulus-category associations.

Another corpus of neuroimaging studies suggests that different brain areas are engaged in different cases of prototype category learning: Implicit prototype learning is reported to be associated with reduced activity in the right occipital cortex (specifically BA-19) when a member of the learned category is processed. When the category prototype is learned explicitly, processing a category member is associated with an increase in neural activity in this region, as well as in the right hippocampus region (Aizenstein et al., 2000; Reber et al., 2003). Zeithamova et al. (2008) presented a somewhat different perspective showing that when prototype learning requires contrasting between two target categories (the A/B task), learning is associated with significantly higher neural activity in the parahippocampus and inferior parietal and orbitofrontal cortices. Hippocampal and parahippocampal activities were also predictive of correct responses on an individual trial basis in this task. On the other hand, performing a task that requires discriminating between members and non-members of a single category (the A/non-A task) is associated with higher activation in the lateral occipital cortex and striatum. Correct categorization in the A/non-A task is associated with significantly higher activation in the left putamen and right anterior hippocampus.

We may deduce from this overview that multiple neural mechanisms in the human brain are involved in category learning. The most common explanations for these findings are that which neural mechanism is engaged during category learning depends on the nature of the learned category structure, the nature of correspondence between observer decision and the provided feedback, and/or whether the category structure is learned implicitly or explicitly. In the current study we present novel findings showing that different neural mechanisms are associated with different forms of induction required for category learning. We suggest that which of these induction tools is engaged for the task of category learning depends on the nature of the information building blocks (specifically, the type of relational information) provided during the learning process, and the computational challenges involved in their processing. To this end, in our experimental task we do not compare the neural correlates of learning one category structure or another, as has been done in previous studies. Instead, we test the neural correlates of learning the same objective category structure while changing the nature of the information provided during the learning process. We argue that studying the underlying neural mechanism of category learning using such an approach may help to further explain and to reconcile some of the apparent inconsistencies across different previous studies discussing the neural basis of category learning.

What kind of mental processes are required, and are potentially available, for any supervised category-learning task? Learning from feedback and observational learning both require processing equivalence constraints – indications of the categorical relations between stimulus examples from which a generalized categorization principle can be learned by comparing these examples (on the role of comparison in category learning see Boroditsky, 2007; Gentner and Namy, 1999; and Spalding and Ross, 1994; see also Goldstone et al., 2010 for a recent review). When we are presented with two stimuli sharing the same label (i.e., members of the same category), we are provided with a *Same-Class Indication* – an indication from which we can identify the features that are shared by category members, as well as the permitted within-category variability (the irrelevant features in which category members may differ). When we are presented with two stimuli that differ in their labels (i.e., members of different categories), we are provided with a *Different-Class Indication* – an indication enabling the identification of diagnostic features that are required for discriminating between different categories (Hammer et al., 2009a).

Same-class indications can also be the product of feedback, signaling a Hit or a Miss – when making a same/different decision in regard to the categorical relation between two members of the same category, not being aware of this fact, *a priori*, the feedback that follows such a decision indicates that these are indeed members of the same category. On the other hand, different-class indications are the product of feedback signaling a Correct-Rejection or False-Alarm – when making a decision in regard to the categorical relation between two members of two different categories, the following feedback indicates that these are members of different categories. The same principle is true in supervised learning scenarios in which a decision is made in regard to the categorical relation between a single exemplar and a given category representation (i.e., A/non-A task) or multiple category representations (e.g., A/B task).

Same-class and different-class indications fundamentally differ in the nature of the computation required for processing them: (1) Same-class indications are always highly informative for learning irrelevant within-category variability and relevant within-category similarities, whereas different-class indications are informative for learning relevant between category differences only in the rare cases when these differences are very few. (2) Same-class indications are transitive, while different-class indications are not. This makes same-class indications even more informative as it also reduces the computational effort in accumulating and binding information provided from multiple observations or learning trials. For these reasons, same-class and different-class indications differ in their usability for learning different category structures (Hammer et al., 2008).

Specifically, same-class indications may be usable both for learning parametric similarity-based representations, such as the one required for prototype or exemplar-based learning, or for learning an explicit rule which is based on induction. On the other hand, different-class indications are effective for rule learning only when they make it possible to evaluate the relevance of very few feature-dimensions in isolation. Different-class indications may also become valuable when simultaneously learning more than a single prototype-based category representation, a scenario in which discrimination capabilities are of greater importance (such as with the case of the A/B category learning task). Yet again, such a use of different-class indications is possible only when the to-be-considered between-category differences are very few. Since different-class indications are not transitive, any task that requires their use will require additional effort for binding information from isolated observations or learning trials (Hammer et al., 2008).

As learning a given category structure using only different-class indications necessitates a different form of computation than the one

Author's personal copy

required for learning the same category structure using only same-class indications, we hypothesize that this may be reflected in the neural correlates associated with each one of these two processes. Accordingly, we tested human participants while they performed complex rule-based category learning tasks, recording their brain activity using functional MRI scanning (whole brain recording). Specifically, we hypothesize that learning a categorization rule from different-class indications is likely to require an induction process, whereas same-class indications are likely to be automatically incorporated by other learning systems, some of which do not require explicit induction (e.g., systems involved in learning a similarity-based category representation). The main motivation of using a complex rule-based category learning task is that in such a task it is relatively easy to formulate the hypothesis space representing the possible ways for categorizing the stimuli prior to the learning process, as well as the way in which this hypothesis space is constrained by the information provided during the learning process (Krawczyl et al., 2005; Holyoak, 2008). This enables quantifying the amount of objective information provided in different learning conditions, and then using this measure as a reference for better evaluating participants' performance (e.g., Hammer et al., 2008). It is important to note at this point that rule-based category learning is believed to be based on explicit hypothesis testing that relies heavily on explicit memory systems in the human brain (DeCaro et al., 2008; Waldron and Ashby, 2001). We argue that this may be only partially true as it may depend on the type of indications (same-class vs. different-class) available during the learning process.

Methods

Outline

When designing the experiment we took the following methodological precautions: (1) We disassociated the two types of indications so that in one category-learning condition participants were trained with only same-class indications while in the other they were trained with only different-class indications. (2) For each category-learning condition we used same-class and different-class indications that were equally informative, making it equally likely to learn the categorization rule if using a strictly optimized induction strategy. (3) For each task, the learning phase was disassociated from the test phase so that participants were not provided with any feedback for their performance during the experiment. (4) Although the task was an observational learning task, we avoided the use of category labels. (5) In the two category-learning conditions, participants were encouraged to attend to both similarities and differences within each compared stimulus pair. (6) To better disassociate the neural correlate of category learning from that of other more generic mental processes, we included a control condition that did not require learning, but in which participants were presented with the same visual input and which required generating similar motor responses as in the category-learning conditions.

Participants

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

Materials

Twelve sets of computer-generated grayscale images of novel "alien creatures" were used as stimuli. Each set was characterized by four binary feature-dimensions that potentially could determine the creatures' categories for the given set. Each set was used for a different category learning task presented in a different experiment block. For each set, categories were predefined as a rule, based on two feature-dimensions only. The background in each image included a low-contrast pattern with blurred and noisy circles or squares scattered in varying patterns. These background shapes become relevant for the control task. Fig. 1a illustrates a few stimuli from one stimulus set.

Procedure for the behavioral task

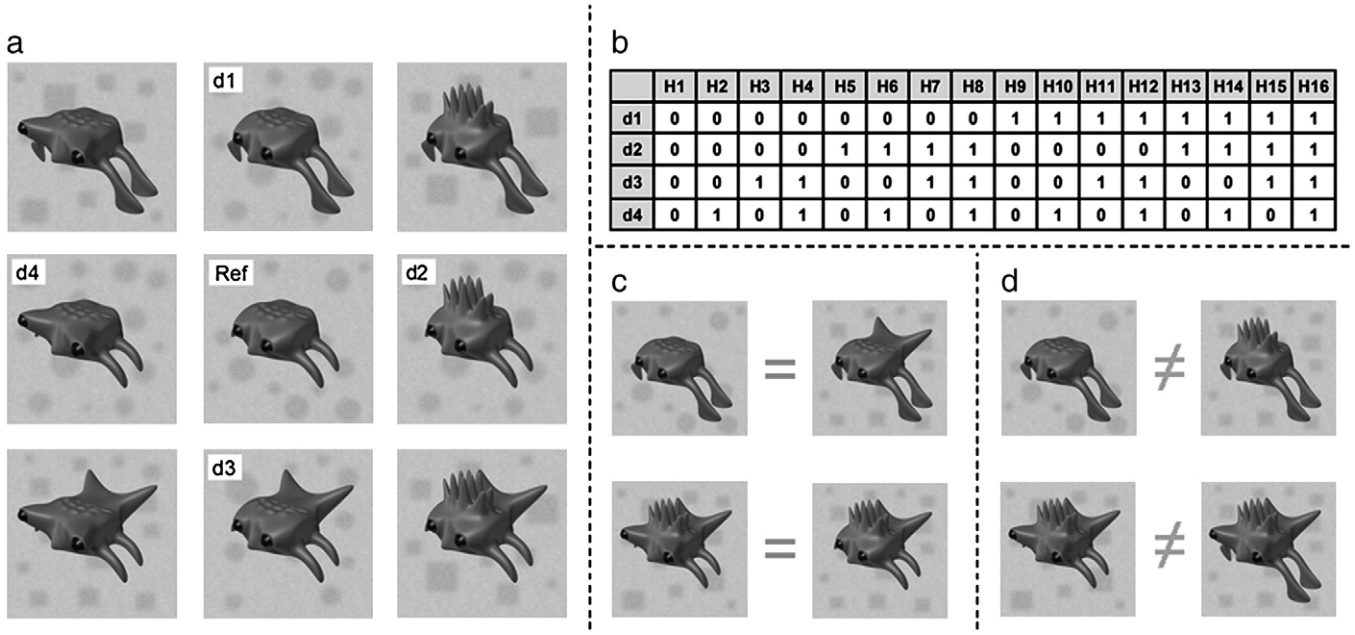
Before the fMRI Scanning session, participants were introduced to the task in a *Stimulus Familiarization* session. This took place outside the fMRI scanner, and participants were presented with stimulus pairs and asked to decide whether the two simultaneously presented creatures are of the same/different type (by pressing the left/right mouse key). No feedback followed this decision so that the predetermined categorization rule could not be learned at this stage. The motivation for conducting this unsupervised discrimination task was to enable participants to become familiar with the stimuli and the potentially informative feature-dimensions prior to the *Scanning Session*.

Since for each stimulus set there are four possibly relevant feature-dimensions, and since the number of task relevant feature-dimensions was unknown to the participants, the hypothesis table in Fig. 1b represents all possible combinations of relevant vs. irrelevant feature-dimensions that could be considered by the participants: Hypothesis 1 (H1) represents a scenario in which none of the varying feature-dimensions is relevant for categorization and H16 a scenario in which all feature-dimensions are relevant, (a conjunction rule based on four feature-dimensions). The other hypotheses represent the other possible alternatives for one, two, or three relevant feature-dimensions. It is important to emphasize that participants were not informed of the number of feature-dimensions in which the stimuli varied, nor the number of feature-dimensions which were with diagnostic value.

After the *Familiarization Session*, participants started the *Scanning Session* in which they were tested in three conditions: Category learning by comparing *Same-Class Exemplars*, category learning by comparing *Different-Class Exemplars*, and a *Control* condition in which the participants did not learn a categorization rule. For each of the category-learning blocks a different stimulus-set was used, requiring learn a new categorization rule. For the control condition we used stimulus sets also used for the same-class and different-class conditions. Each experimental condition (same-class, different-class, and control) was repeated six times during the scanning session, with blocks from the three conditions being presented in random order. At the beginning of each block, a slide indicating the specific task (saying – "same-type clues", "different-type clues", or "background shape") was presented to the participant for 3.5 s (followed by a 0.5 s blank screen). In the two category-learning conditions, this slide was followed by a learning phase, during which the participant could learn a categorization rule by comparing pairs of creatures identified to be members of the same-category or of two different categories. Same-class and different-class pairs were selected for the learning phase to provide the participants with the same amount of information in the two learning conditions. In each category-learning condition, the learning phase was followed by a test phase with eight test trials. Participants were not provided with any feedback for their performance (at any point during the experiment).

In the same-class condition participants could learn the categorization rule from two same-class indications, each providing one bit of information by eliminating half of the possible hypotheses; (see Fig. 1). In each of the two learning trials, a pair of creatures were presented with an equal sign ("=") presented between them, indicating that these paired creatures are of the same type. Fig. 1c illustrates two same-class indications: From the upper indication

Author's personal copy



	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16
d1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
d2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
d3	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
d4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Fig. 1. Stimuli design. (a) Example of stimuli taken from one of the 12 stimulus sets used in the experiment. Creatures in each stimulus set vary in four feature-dimensions (for the current example, d1: legs; d2: spikes; d3: tail; d4: eye protrusion). For each set we created 16 stimuli, representing all combinations for the four binary dimensions). (b) Hypothesis table illustrating all possible combinations for relevant ("1") or irrelevant ("0") feature-dimensions. (c) Examples of same-class pairs indicating d3 (tail; upper pair) and d4 (eye protrusion; lower pair) as irrelevant. (d) An example of different-class pairs indicating d2 (spikes) and d1 (legs) as relevant feature-dimensions.

participants could learn that same-category creatures may differ in their tails (marked as d3 in Fig. 1a). Thus, tails are not relevant for categorization since the within-category variability in this feature-dimension is equal to the overall observed variability in this stimulus set. From the lower same-class indication, participants could learn that same-category creatures may differ in eye protrusion (marked as d4 in Fig. 1a) indicating that this feature-dimension is also irrelevant. Each same-class creature pair was presented separately for 5 seconds (followed by a blank screen for one second). Together, the two same-class indications provide two bits of information leaving only the four possibilities in which d3 and d4 are both irrelevant for the categorization rule (marked by "0" in the hypothesis table). This leaves H1, H5, H9 or H13 as possible hypotheses for the categorization rule.

In the *Different-Class* condition participants could learn the categorization rule from two different-class indications, each providing one bit of information (Fig. 1d). Here, pairs of "creatures" were presented with the not-equal sign ("≠"), indicating that these creatures are from different types. From the upper different-class indication of the figure, participants could learn that if creatures differ in their spikes (marked as d2 in Fig. 1a) they are necessarily from different types. That is, spike length is relevant for categorization since it is a diagnostic feature. From the lower different-class indication participants could learn that difference in creature legs (marked as d1 in Fig. 1a) is also relevant for discriminating between categories. Together, the two different-class indications provide two bits of information leaving only the four possibilities in which d1 and d2 are both relevant (marked by "1" in the hypothesis table). This includes H13, H14, H15 or H16 as the possibly relevant hypotheses. Note that in both learning conditions – using either only same-class or only different-class indications – the information is equal and efficient learning will result with an equal number of alternative hypotheses. The hypothesis by which participants' performance was evaluated in this example is H13, stating that d1 and d2 are the only two relevant feature-dimensions. H13 is the only hypothesis consistent with both the same-class and the different-class indications provided (note that participants never saw the same stimulus set for both learning conditions). Describing it in other words, for the current stimulus set example, the objective category structure that complies with both the same-class and different-class indications described above can be verbalized as a conjunction rule of one type of spikes and one type of legs (whereas eye protrusion and tail type are irrelevant for this categorization task).

Immediately after being trained with either the same-class or the different-class indications, participants were tested on the categorization rule that they had just learned. In each of the eight test trials that followed the test phase, two creatures were presented with no indication for their categorical relation. In each of the test trials, the two presented creatures were identical in exactly two out of the four possible feature-dimensions, and differed in the other two. Thus, the amount of similarity vs. dissimilarity with respect to the four varying features was balanced, reducing possible response bias. In half of the test trials the paired creatures differed in the two irrelevant feature-dimensions and were identical in the two relevant feature-dimensions (i.e., same-class creatures). In the other half of the test trials the creatures differed in one relevant feature-dimension (i.e., different-class creatures) and one of the irrelevant feature-dimension. Participants made a same/different decision (pressing the appropriate key of the response box) according to what they had learned during the preceding learning phase (but did not receive feedback for their decisions). Each test trial lasted 3 s (2.8 s for stimulus presentation and response, and 0.2 s inter-trial-interval with a blank screen). Between blocks there was a 20 s blank screen rest period.

Control condition blocks were similar in structure, stimulus type and duration. However, here participants were not asked to categorize the presented creatures but to decide whether the low-contrast background shapes in the two stimuli are the same (square-square or circle-circle) 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 were task irrelevant. To insure that this task is sufficiently demanding, the stimuli were designed so that the background shape distribution pattern varied between stimuli (thus, it could not be predicted), they were blurred and noisy and with low-contrast (see examples in Figs. 1 and 2). Thus, although the control condition did not involve category learning, participants were presented with similar visual input, and had to

Author's personal copy

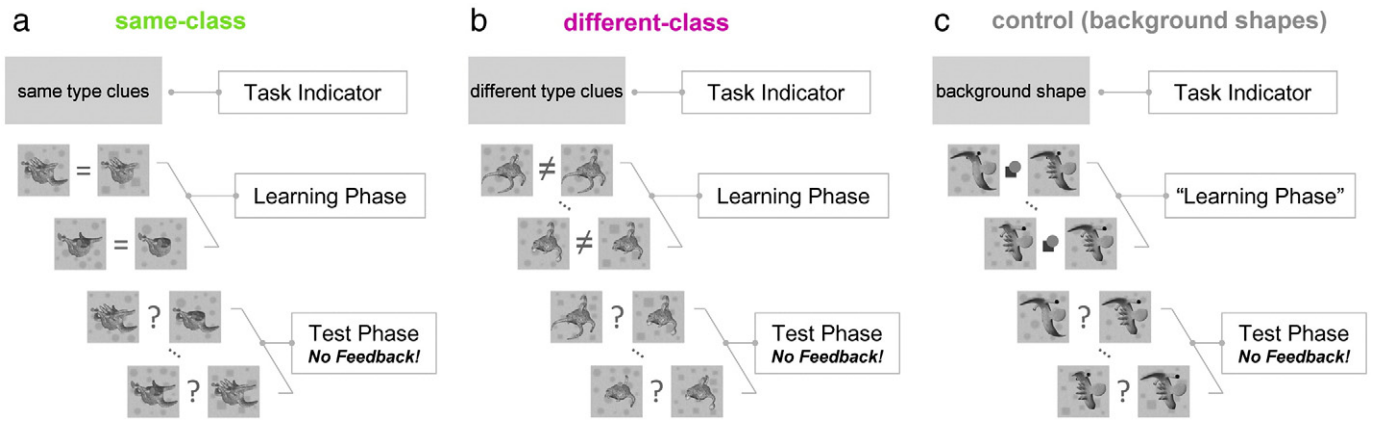


Fig. 2. Outline of a block design in each experimental condition. (a) Illustration of one category learning block of the same-class indications condition. The block started with a slide indicating to the participant the block type (overall duration 4 seconds). This was followed by the learning phase (12 s) in which two same-class exemplar pairs were presented with the "=" sign, indicating for the participants that each two stimulatingly presented creatures are members of the same category. During the following 8-trial test phase (24 s), participants had to decide whether the presented pair of creatures was from the same type or different types (pressing the appropriate response-box button), receiving *no feedback*. (b) Illustration of one block of the category learning from different-class indications condition, in which two different-class exemplar pairs were presented with the "≠" sign, indicating for the participants that each two stimulatingly presented creatures are members of two different categories. (c) Illustration of one block of the control condition; here, during the "learning phase," no information for the categorical relation between the paired stimuli was provided, so that no learning could take place. In the control condition, instead of learning a categorization rule, participants related to the low contrast background shapes.

produce similar motor output as in the two category learning conditions. This enables us to differentiate neural activity associated with category learning from that accompanying the processing of visual input, general visual attention and the production of motor output (eye saccades and key pressing). See Fig. 2 for an illustration of the general block design in each condition.

Scanning procedure

For stimulus presentation and participant response recording, we used Presentation software (www.neurobs.com) running on a standard PC computer. Stimuli were back-projected onto a screen which could be viewed via a mirror mounted on the head coil. The distance between the participant's eyes and the screen was 59 cm. The screen was 325×260 mm, which is appropriate for an angle of about ±15°. In each experimental trial, the pair of stimuli was simultaneously presented at the center of the screen. Each stimulus occupied 10×10 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 response box (for same/different responses, respectively).

Measurements were carried out on a 3 T scanner (Siemens Trio, Erlangen, Germany) equipped with an eight channel head coil. A 3D anatomical data set of the participant's brain (192 slices of 1 mm each) was obtained before the functional MRI measurement. Additionally, we acquired an *Inversion-Recovery-Echo-Planar-Imaging* (IR-EPI) scan with the identical geometry as in the functional measurement. 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, 64×64; field of view, 19.2 cm×19.2 cm; 33 slices of 3 mm thickness with 0.3 mm gaps). During the scans, participants wore earplugs and their head was fixed with a cushion.

Procedure for scanning data analysis

The functional data were analyzed with BrainVoyager™QX 1.8.6 software (Brain Innovation, Maastricht, Netherlands). A standard sequence of preprocessing steps, including 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 and spatially smoothed with a Gaussian filter (FWHM = 4 mm). For the fixed-effects GLM-analysis, we defined the following predictors for each 1 minute block: task indicating slide (0–3.5 s), first (same/different or control) indication (3.5–9.5 s), second (same/different or control) indication (9.5–15.5 s), test phase (15.5–39.0 s), and rest (39.0–60 s). These were convolved with the two-gamma hemodynamic response functions using the default parameters implemented in BrainVoyager™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 using a significance level of $p < 0.05$ (FDR-corrected). The contrasts between each of the two category learning conditions and the control condition are illustrated in Supplementary Fig. 1. This analysis step was done as an initial step for all further analyses (see Results) that reveal effects specific for each of the category learning conditions, using a significance level of $p < 0.05$ (Bonferoni corrected) and a minimum cluster size of 100 mm^3.

Results

Behavioral performance

We measured category learning efficiency as participant performance in the 8 test trials in each block. For each participant we averaged performance from the six blocks of each condition. The main performance measure we used is the non-parametric sensitivity measure A′ (Grier, 1971), calculated from the Hit rate (correctly identifying two creatures as belonging to the same category) and False-Alarm rate (FA; incorrectly identifying two creatures as belonging to the same category) of each participant in each condition separately. $A' = 0.5$ represents chance performance, $A' = 1$ perfect performance; $0 \leq A' < 0.5$ represents response confusion (particularly likely when a participant produces many missed trials).

One-sample *t*-tests showed that participant performance in all three conditions was significantly better than chance ($A' = 0.5$): Learning from different-class indications ($A' = 0.84 \pm 0.17$; Mean ± SD), $t(13) = 7.23$, $p < 0.001$; Learning from same-class indications ($A' = 0.64 \pm 0.17$), $t(13) = 3.10$, $p < 0.01$; Control (background shapes; $A' = 0.96 \pm 0.03$), $t(13) = 50.21$, $p < 0.001$. Importantly, performance in the learning from different-class condition was significantly better than that in the same-class condition, $t(13) = 3.52$, $p < 0.005$. As shown in Fig. 3, this difference in sensitivity between the two category learning conditions reflects the somewhat higher Hit

Author's personal copy

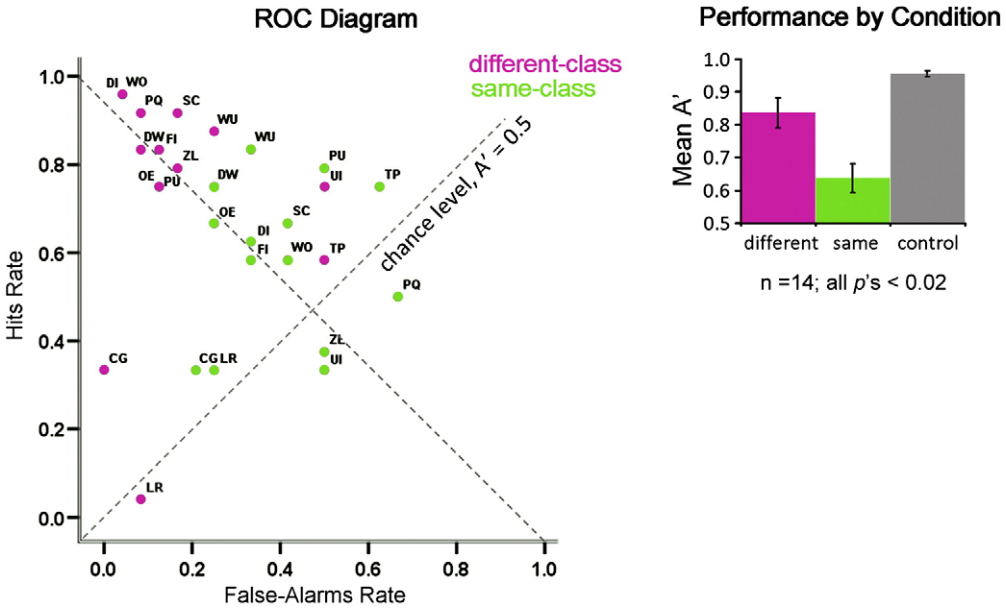


Fig. 3. Behavioral performances. (Left) each participant (marked with his or her initials) is represented twice on a Receiver Operation Characteristic diagram, once for the same-class condition (green) and once for the different-class condition (pink). (Right) mean (and standard error of the mean) sensitivity (A′) in the three experimental conditions.

rate, $t(13)=2.51$, $p<0.05$, but mainly the lower FA rate, $t(13)=-5.92$, $p<0.001$, when learning from different-class indications (Hit rate $=0.74\pm0.26$; FA rate $=0.16\pm0.16$) compared to same-class indications (Hit rate $=0.58\pm0.18$; FA rate $=0.40\pm0.14$). See Discussion for a comment on the possible source of this effect.

Neural correlate of learning from different-class indications

In order to characterize the neural correlate of category learning from different-class indications we contrasted the neural responses in the two category learning conditions during the learning phase (different-class higher than same-class), as shown in Fig. 4. This reveals significantly higher neural activity in an assembly of cortical and subcortical brain areas including the left lingual gyrus (ventral BA-19; Fig. 4a), left and right putamen and the left caudate (Fig. 4b, c), and the right hippocampus (Fig. 4d, showing the contrast in the hippocampus for the test phase). Fig. 4e shows the nature of the differences in the neural activity in the two category learning conditions, also with respect to the control condition. Note that, excluding the neural activity in the left caudate, the neural activity in the other listed brain areas is no higher than the neural activity in the control condition, which requires no category learning. Moreover, for the most part, the significant effects in these brain areas seem to be exclusively associated with learning: As can be seen in Fig. 4f, most of these effects dramatically decrease in the subsequent test phase. Although the neural activity in the different-class condition remains significantly higher in the right and left putamen, the effect size is reduced. The only effect to become even more substantial during the test phase is the one observed in the right hippocampus. It is important to note that when evaluating the pattern of BOLD responses during the test phase, one should remember that this is likely to be affected by the pattern of BOLD responses observed in the preceding learning phase, so that even the reported effect seen in the hippocampus might be due to learning stage differences, not test stage differences.

Looking at the greater neural activity in the learning phase of the learning from different-class indications condition and recalling that performance, too, was superior in this case, there is an apparent overall positive correlation between learning stage neural activity (in the dorsal striatum, hippocampus and lingual gyrus) and participant performance in the test phase (across the two category learning conditions). Thus, a possible conclusion would be that the level of neural activity in these brain areas is not truly associated with the type of indications from which the categorization rule is learned (i.e., same- vs. different-class indications), but rather the overall level of neural activity in the learning phase is simply correlated with the overall expected level of performance in the following test phase. Given the temporal order, one might even claim that neural activity in the learning phase predicts (or even associated with the cause of) participant competence in the following categorization task (test phase).

Although the above conclusion might seem compelling, we argue against it. The basis for negating this conclusion is that when we look at the across subjects correlation between activity in these areas and behavioral performance, for each one of the two category learning conditions separately, we find no correlation at all. To test this correlation, we average the BOLD signal across the three relevant dorsal striatum areas (left and right putamen and left caudate; marked as clusters 1, 2 and 3 in Fig. 4). This step is justified both by what is known about the anatomical and functional relations among these brain structures (Bar-Gad et al., 2003), and by the analysis presented in the Supplementary data, showing that the neural activity in these three brain areas is highly correlated in the tasks tested here (during both the learning and the test phases). The correlations among the activation in these three sub-structures in the dorsal striatum with activation in the left lingual gyrus and right hippocampus are also provided as a reference (see Supplementary Tables 1–4). This analysis shows that the correlations among the neural activity in left lingual gyrus, right hippocampus and sub-structures in the dorsal striatum strongly depend on experimental condition (same- vs. different-class) and task phase (learning vs. test).

Fig. 5a shows the correlation between the neural activity (Beta) in the dorsal striatum in the learning phase and the level of performance (*A′*) in the test phase for each experimental condition. Within each condition there is no significant correlation between the two parameters: Different-class condition, $r(14)=0.02$, $p=0.95$; Same-class condition, $r(14)=-0.19$, $p=0.52$. This is also the case when looking at correlations between the neural activity in the dorsal striatum in the test phase and the level of performance for each experimental condition separately: Different-class condition, $r(14)=$

Author's personal copy

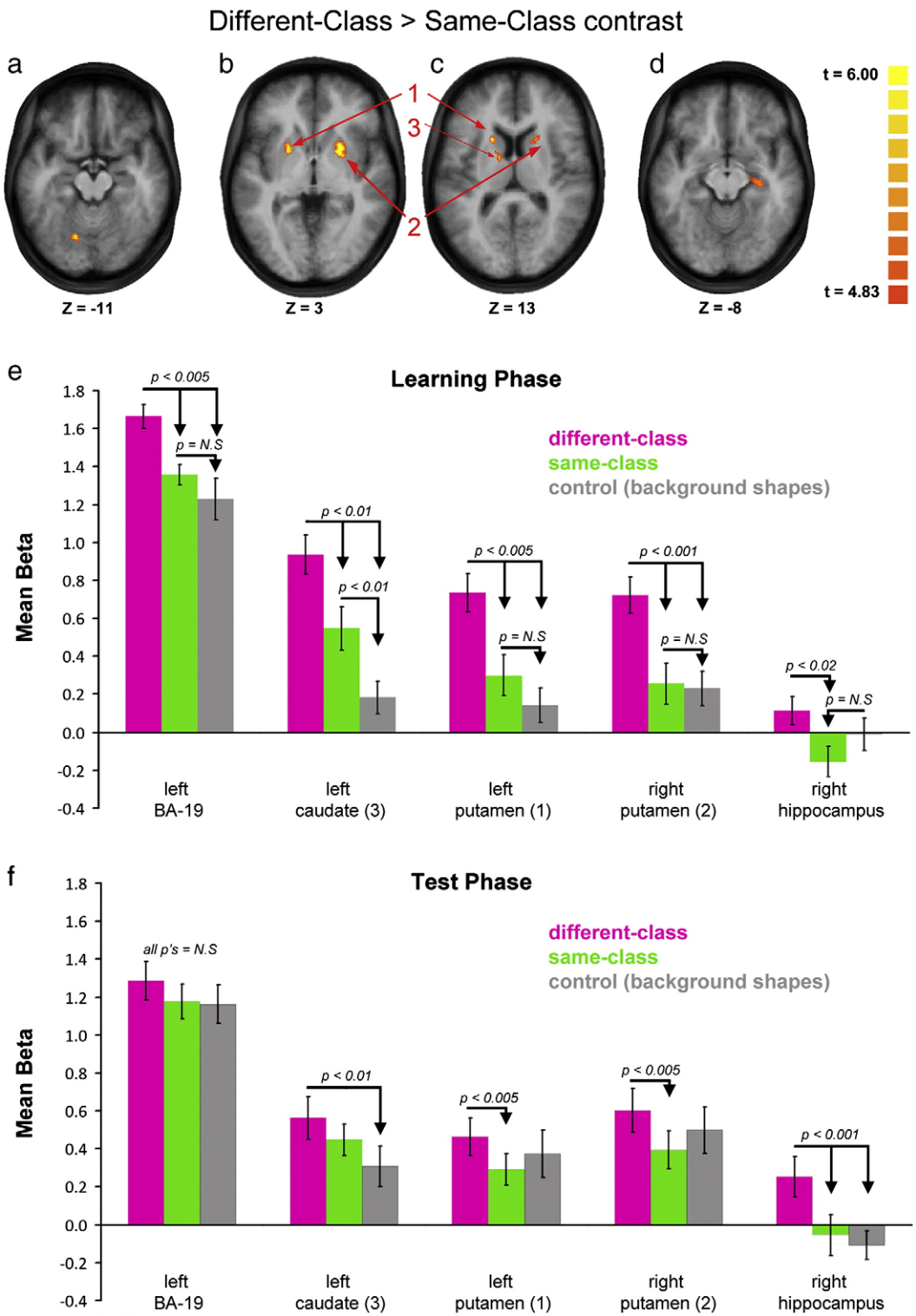


Fig. 4. The neural correlate of learning from different-class indications. Neural activity in brain areas determined by the different-class>same-class contrast for the learning phase (a, b, and c) and the test phase (d). (a) Activation in left lingual gyrus (ventral BA-19). (b) Bilateral activation in the putamen. (c) Bilateral activation in the putamen (continues the voxel clusters as in panel b) and in the left caudate. (d) Activation in the right hippocampus. (e, f) Mean Beta (and standard error of the mean) for each of the functional ROIs, for each experimental condition, in the learning phase (e) and test phase (f); the latter provides a reference demonstrating that differential activation in the two category learning conditions is mainly restricted to the learning phase.

0.34, $p = 0.23$; Same-class condition, $r(14) = 0.31$, $p = 0.28$. It is important to note that these correlations are not significant although there is substantial variability in participants' performance and in the neural activity within each of the two category learning conditions.

In addition, we considered the possibility that the relatively low neural activity in the same-class conditions results from more participants being short of attention or confused, specifically when performing this category learning condition. Accordingly, we excluded five participants for whom performance was poor in either one of the category learning conditions ($A' < 0.5$, associated with response confusion, or relatively many trials in which the participant did not responded on time), or that their dorsal striatum neural activity

Author's personal copy

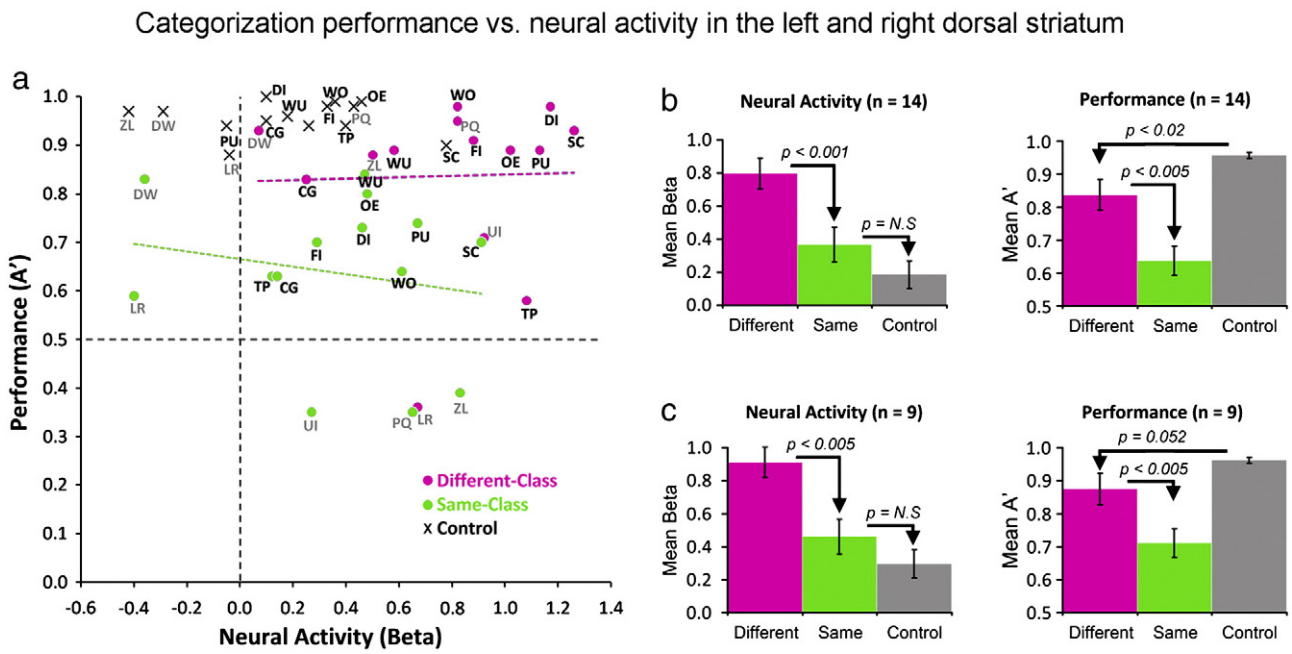


Fig. 5. Behavioral performances vs. neural correlate (learning from different-class indications condition) in the left and right dorsal striatum. (a) Scatter plot showing relation between neural activity in the left and right dorsal striatum (combined) and performance in each of the three experimental conditions. Dashed pink and green lines are regression lines of different- and same-class conditions, respectively (see text). (b) Mean neural activity and standard error of the mean (left) and performance (right) averaged over 14 participants. (c) Mean neural activity and standard error of the mean (left) and performance (right) for best 9 performers.

during the learning phase was lower than during the between-block intervals (which might be associated with not paying attention to the stimuli). Fig. 5 presents mean performance and level of neural activity in the dorsal striatum of all participants (5b) or after excluding the five poorest performers (5c). Apparently, excluding these poor performers (who indeed performed poorly specifically in the same-class condition) has no significant effect on the overall pattern of neural activity though it brings the performance level in the three (two category learning and the control) conditions closer to equality.

The lack of significant correlation between neural activity in the learning phase and performance in the test phase, within each one of the two category learning conditions separately, suggests that it is unlikely that the significant difference between the two conditions in these two parameters results from some simple interdependence between them, which has nothing to do with the nature of the provided information. Instead, it is more likely that the type of indications (same-class vs. different-class) provided during the learning phase *directly* affects both the brain areas engaged in the task of category learning and the outcome of the learning process (Table 1).

To lend further support to this latter conclusion, we performed a regression analysis to test whether activity level adds predictive power beyond the learning-condition effect. We converted the categorical variable of category learning condition to an interval-scale dummy variable (labeling the same-class condition with the value "0" and the different-class condition with the value "1"). Next, we calculated the linear regression between the three factors referring to the category learning conditions and the neural activity as independent variables, and the participants' performance as the dependent variable. The regression analysis (using the "enter" method) shows that participant performance is predicted from the learning condition, $\beta(18) = 0.57$, $p < 0.04$, and that adding the neural activity during the learning phase to the regression analysis has no additional explanatory power, $\beta(18) = 0.13$, $p = 0.61$. This suggests that differences in neural activity that cannot be attributed to category learning condition have no direct effect on performance. Table 2 shows the Pearson correlations among the three parameters.

Neural correlate of learning from same-class indications

Finally, we note that the neural correlate of category learning from same-class indications (same-class > different-class contrast) shows a significant effect in a cluster of only 88 voxels in the right inferior temporal gyrus (BA-37), and two clusters in the right angular gyrus (BA-39), with 69 and 49 voxels, respectively (all at a significance level of $t = 3$; see Fig. 6 and Table 3). Since these effects are significant but limited to relatively small clusters of voxels, we consider them with caution. Nevertheless, we report these effects since the right and left BA-37 and BA-39 cortices have been reported by others to be involved

Table 1
Talairach coordinates for ROIs determined by the different-class > same-class contrast in the learning phase (the test phase for the right hippocampus).

	x	*Y*	*z*	No. of voxels	Beta	*t*
Left ventral BA-19	−14	−66	−8	392	1.62	6.6
Left caudate	−12	3	11	270	0.77	6.4
Left putamen	−19	11	7	572	0.87	7.1
Right putamen	23	7	5	1446	0.93	7.7
Right hippocampus	26	−19	−7	204	0.31	7.0

Table 2
The binominal Pearson correlations between the experimental condition (using the dummy values same-class = "0"; different-class = "1"), the neural activity in the left and right dorsal striatum (Beta values), and the participants performance (*A*′ values).

	Category learning condition	Participants performance
Neural activity in the dorsal striatum	$r = 0.63$ $p < 0.005$	$r = 0.49$ $p < 0.05$
Category learning condition		$r = 0.66$ $p < 0.005$

Linear regression analysis for these three factors suggests that the significant correlation between neural activity and in the dorsal striatum and the participant performance is the byproduct of the other two correlations presented in this table (see text).

Author's personal copy

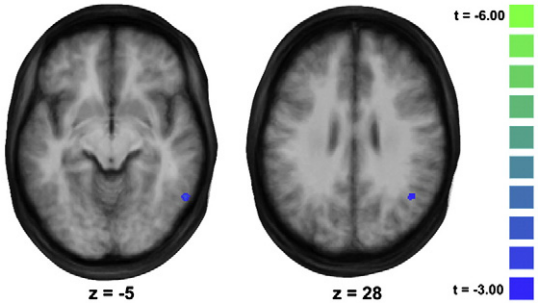


Fig. 6. The neural correlate of learning from same-class indications. Neural activity in brain areas determined by the same-class>different-class contrast for the learning phase. (Left) Activation in the right inferior temporal gyrus (BA-37). (Right) Activation in the larger cluster in the right angular gyrus (BA-39).

in prototype-based category learning (Reber et al., 2003; see Kéri, 2003).

Discussion

Our findings show that learning a categorization rule from different-class indications is associated with significantly higher neural activity in the left lingual gyrus (ventral BA-19), left and right dorsal striatum, and right hippocampus as compared to the condition where the same objective category structure is learned by processing same-class indications. Moreover, when processing same-class indications, the neural activity in these brain areas, for the most part, is no higher than in a condition that did not require any category learning but in which the same visual input was presented and participants produced a similar motor output. The reverse contrast (same-class>different-class) shows that the task of category learning from same-class indications is associated with marginally higher neural activity in the right inferior temporal gyrus and right angular gyrus, brain areas involved in high-level visual processing. This overall pattern of neural activity is striking – contrasting the two comparison processes reveals significant effects in brain areas that have been reported by others to be involved in learning different category structures.

These findings suggest that different neural mechanisms are employed for executing two different fundamental comparison-based induction processes – namely, learning a categorization principle by comparing different-class exemplars versus learning by comparing same-class exemplars. We argue that since the processes of learning different types of category structure (e.g., rule-based, prototype-based, or information-integration-based representation) may differ in their dependence on one or both of these two types of induction-based comparison processes, the relative weights of the neural mechanisms associated with executing these two processes may differ when learning different category representations.

Separate cerebral structures reflecting different computations

Why does not the neural circuitry involved in category learning from different-class indications simply overlap the one required for processing same-class indications? What is the possible source or purpose for such an anatomical and functional disassociation? As we suggested in the Introduction, the answer may be related to the fact that the process of learning a categorization rule by comparing different-class exemplars differs fundamentally from that of learning such a rule by comparing same-class exemplars. This fundamental difference relates to a number of elements of category learning, which we review in the following paragraphs. These include the nature and quantity of acquired information, and the degree to which information must be integrated over multiple comparisons.

The first obvious difference is that comparing different-class exemplars is most effective for directly identifying *diagnostic* features, whereas comparing same-class exemplars is most effective for identifying within-category *common* features. Comparing same-class exemplars may be even more effective for excluding irrelevant feature-dimensions by mapping permitted within-category variability. In fact, same-class indications may be useful for a variety of fundamental visual system processes such as establishing an invariant representation of an object by aligning and comparing its multiple representations (Bar-Hillel et al., 2005).

In addition, the quantity of information provided by a same-class indication is typically far greater than that provided by a different-class indication: When we are informed that two exemplars which differ in many visible feature-dimensions are members of the same category, we can conclude that these feature-dimensions are irrelevant for categorization. As the number of task-irrelevant feature-dimensions increases, the information gain from comparing same-class exemplars also increases. This is not the case when comparing different-class exemplars: When two different-class exemplars differ in many feature-dimensions, it becomes impossible to deduce which of the differences between the two has diagnostic value for discriminating between the represented categories. In fact, only when there is a single difference may one be able to decisively identify it as a diagnostic feature. Thus, learning a complex rule or any category structure based on multiple feature-dimensions, using mainly different-class indications, will be more demanding in the need to incorporate information from multiple comparisons. Given the fact that same-class indications are transitive, but different-class indications are not, makes this difference in computational effort even more prominent. In the experimental task we used here, we deliberately provided participants with same-class and different-class exemplar pairs that always differed in a single feature-dimension (irrelevant or relevant, respectively). Hence, in both conditions we used indications providing a single bit of information, making it similarly possible to learn the categorization rule, when using an optimal induction strategy (See Hammer et al., 2008 for a formal proof for this statement). Nevertheless, we found that the neural correlate of these two comparison processes is still different even in this case, suggesting that this difference represents an essential difference in computation and not quantitative differences.

Behavioral implications of the different category learning mechanisms

Our finding that different cortical structures are involved in these two types of category learning is consistent not only with the different natures of their respective computational challenges, but also with a series of differences that we have reported between the ensuing performance strategies, levels and development, as reviewed in the following paragraphs.

In previous studies we found that the objective computational differences between same-class and different-class indications have a striking effect on human behavior even when tested under conditions in which the advantages of learning from same-class indications are eliminated: Both children and adults are capable of learning a categorization rule when provided with very few same-class indications, though this always ends with reasonably good yet “suboptimal” performance. Proficiency of learning categorization rules when

Table 3
Talairach coordinates for the ROIs determined by the same-class>different-class contrast in the learning phase.

	x	*y*	*z*	No. of voxels
Right BA-37	54	55	−5	88
Right BA-39 (I)	40	−54	29	69
Right BA-39 (II)	41	−58	38	49

Author's personal copy

provided with similarly-informative different-class indications develops only at late childhood (Hammer et al., 2009b). In fact, learning from (the rarely) informative different-class indications is not always intuitively employed even by many adults, though when employed it leads to superior categorization performance, compared to learning solely from same-class indications (Hammer et al., 2009a,b; and current Results).

Learning category structure from poorly informative different-class indications may be possible only by applying a statistical learning strategy. One would need to detect regularities, noting that some combinations of between-category differences co-occur more often than others, suggesting that these combinations may carry diagnostic value. This type of learning requires many learning trials, and the current experiment was not designed to test such potential learning mechanism (see Turk-Browne et al., 2008, for a possible cognitive mechanism involved in statistical learning of a multidimensional representation, and Shohamy et al., 2004, Turk-Browne et al., 2009, for a possible neural mechanism for detecting category regularities).

Another difference between these strategies may be that learning a category structure based on few relevant feature-dimensions, using informative different-class indications, is more likely to involve the declarative learning system. Though we do not provide direct behavioral evidence for this hypothesis in the current study, this idea is supported by recent behavioral findings: Providing verbal directions to adult participants, encouraging them to use an explicit induction-based learning strategy, significantly boosts performance when they are learning a categorization rule from different-class indications, but not when they learn the same category structure from same-class indications (Hammer et al., 2009a).

On the other hand, same-class indications may invite people to use alternative strategies: (1) Using induction-based learning for excluding irrelevant feature-dimensions (differences between the compared same-class exemplars). After filtering out irrelevant within-category differences, the categorization rule can be based on the remaining salient similarities between the compared same-class exemplars. Since this induction-based strategy does not enable direct identification of relevant feature-dimensions, its implementation may result in participants' missing less salient yet important within-category similarities which can be also used as between-category differences. (2) Learning a similarity-based category representation, such as an exemplar-like representation that maps permitted within-category variability, a prototype-like representation, based on constructing a prototypical class element, or a combination of the two. This potentially non-declarative learning strategy can be sufficient for learning a representation that will support better-than-chance performance even when using non-transitive same-class indications as in the current experiment (see Hammer et al., 2007, for a computer simulation demonstrating the feasibility and performance of such a classifier). These strategies for processing same-class indications are likely to result in suboptimal performance though the objective amount of information provided by same-class indications is typically higher than that provided by different-class indications. This result is consistent with the current behavioral findings showing relatively constrained performance when learning the categorization rule solely from same-class indications. It is important to note here that providing the participants with an opportunity to explore the to-be-categorized stimuli, prior to the learning from comparison phase, helps improving learning from same-class indications more than it helps improving learning from different-class indications (see Hammer et al., 2009a, 2009b for further discussion).

Roles of different cerebral structures in comparison-based learning and induction

Returning to our functional imaging results, what may be the roles of the specific areas that we found activated when participants perform category learning tasks? We found that learning from different-class indications is associated with the dorsal striatum and the hippocampus. These areas may be involved in identification of diagnostic features and creation of associations among them, processes linked to optimal use of different-class indications. Yet, further study is required in order to determine the exact role of these brain structures in category learning. The significantly higher neural activity in the left lingual gyrus (BA-19), associated with processing different-class indications, may suggest that learning which visual features have diagnostic value also involves mid- to high-level visual processing areas. This might be the outcome of directed attention mechanism aimed to highlight specific visual features with mid-level complexity (Hochstein and Ahissar, 2002), or the process of initiating changes in tuning properties to objects or specific object features (Jiang et al., 2007). The dorsal striatum and hippocampus may interact in order to form a category representation by creating proper associations between the task's relevant visually perceived features represented in the lingual gyrus (see Shohamy and Wagner, 2008, and Kurmaran et al., 2009, for related ideas regarding the role of the hippocampus in associating learning events and possible characteristics of this mechanism in decision making and category learning).

Our results suggest that as category learning becomes more heavily dependent on information provided by different-class indications, the dorsal striatum and hippocampus will become more dominant in the learning process. This may be the case in any category learning task that is best performed by identifying very few diagnostic features, as it is the case in rule-based category learning tasks such as the one tested here. This process may also become significant during final stages of learning other, more complex category structures, after most of the irrelevant within-category variability, and some of within-category feature-dimension dependences have been acquired (either from information provided by same-class indications, or based on unsupervised statistical learning).

Turning to the neural correlates of comparing same-class exemplars, this process includes the right inferior temporal gyrus and right angular gyrus (high-level visual areas). In addition, processing same-class indications is associated with a significant correlation between the neural activity in the left lingual gyrus (mid-level visual area) during the learning phase, and participants' performance level in the test phase (see Supplementary Table 4). This suggests that learning to categorize visually perceived objects on the basis of within-category similarities and differences involves higher demands from visual areas that process complex objects and visual features. This also seems to reduce the amount of subcortical structure intervention that may be involved in attention modulation and associative learning. As the behavioral findings show, this learning process results in a relatively simplified representation of the learned category principle. It is possible that this involves creation of a similarity-based category representation instead of one which is based on explicit induction strategy (as seems to be the case for learning based on different-class indications).

Although we only directly tested rule-based category learning task, a case in which quantifying and equalizing the information provided by same-class and different-class indications is relatively easy, our findings suggest that when studying the neural basis of category learning, one must take into consideration the quantitative and qualitative computational differences between these two comparison-based learning processes. We argue that this principle is not limited to rule-learning, and that similar functional disassociation may be observed in the context of other tasks as the relative weights of between-category discrimination and within-category generalization changes (see for example Zeithamova et al., 2008, for potentially relevant findings when contrasting A/non-A with A/B prototype learning tasks). Given the current experimental design, in which participants learned the categorization rule without any feedback for their performance and in the absent of category labels, it seems that

Author's personal copy

the hippocampus and dorsal striatum are not only associated with reward processing or in creating associations between category representations and category members. We suggest, instead, that these brain structures are also involved in other forms of computation that are specifically required for learning a generalized categorization rule by detecting diagnostic features and creating associations between them (see Cincotta and Seger, 2007, for a relevant discussion).

Importantly, the current findings suggest that two different neural mechanisms are involved in facing the two most basic comparison-based learning processes needed for category learning: One mechanism, associated with computing same-class indications, enables ignoring irrelevant dissimilarities and identifying the most relevant within-category similarities. The second mechanism, which is associated with computing different-class indications, enables purposely identifying diagnostic features, which best discriminate objects from different categories. The current findings, together with findings from previous developmental studies, suggest that learning to incorporate these two systems may underlie our capacity to shift from categorization strategies driven mostly by low-level bottom-up factors, to richer, more complex and purpose-oriented representations of objects and events, which better fit our needs (Diesendruck et al., 2003; Hammer and Diesendruck, 2005; Hammer et al., 2009b; Sloutsky, in press).

Acknowledgments

This study was supported by grants from the Israel Science Foundation, the US-Israel Binational Science Foundation, and the European Union under the DIRAC integrated project IST-027787.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2010.03.080.

References

Aizenstein, H.J., MacDonald, A.W., Stenger, V.A., Nebes, R.D., Larson, J.K., Ursu, S., Carter, S.C., 2000. Complementary category learning systems identified using event-related functional MRI. J. Cogn. Neurosci. 12, 977–987.

Ashby, F.G., Gott, R.E., 1988. Decision rules in the perception and categorization of multidimensional stimuli. J. Exp. Psychol. Learn. Mem. Cogn. 14, 33–53.

Ashby, F.G., O'Brien, J.B., 2005. Category learning and multiple memory systems. Trends Cogn. Sci. 9 (2), 83–89.

Ashby, F.G., Noble, S., Filoteo, J.V., Waldron, E.M., Ell, S.W., 2003. Category learning deficits in Parkinson's disease. Neuropsychology 17, 115–124.

Bar-Gad, I., Morris, G., Bergman, H., 2003. Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. Prog. Neurobiol. 71, 439–473.

Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D., 2005. Learning a Mahalanobis metric from equivalence constraints. J. Mach. Learn. Res. 6, 937–965.

Boroditsky, L., 2007. Comparison and the development of knowledge. Cogn. 102 (1), 118–128.

Cincotta, C.M., Seger, C.A., 2007. Dissociation between striatal regions while learning to categorize via feedback and via observation. J. Cogn. Neurosci. 19 (2), 249–265.

DeCaro, M.S., Thomas, R.D., Beilock, S.L., 2008. Individual differences in category learning: Sometimes less working memory capacity is better than more. Cognition 107 (1), 284–294.

Diesendruck, G., Hammer, R., Catz, O., 2003. Mapping the similarity space of children and adults' artifact categories. Cogn. Dev. 118 (2), 217–231.

Filoteo, J.V., Maddox, W.T., Ing, A.D., Zizak, V., Song, D.D., 2005a. The impact of irrelevant dimensional variation on rule-based category learning in patients with Parkinson's disease. J. Int. Neuropsychol. Soc. 11, 503–513.

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

Gentner, D., Namy, L.L., 1999. Comparison in the development of categories. Cogn. Dev. 14, 487–513.

Goldstone R.L., Day S., and Son J.Y., 2010. Comparison. In B. Glatzeder, V. Goel, and A. Von Müller (Eds.) On Thinking: Volume II. Towards a Theory of Thinking. Heidelberg, Germany. Springer Verlag GmbH., pp. 103–122.

Grier, J.B., 1971. Nonparametric indexes for sensitivity and bias: Computing formulas. Psychol. Bull. 75, 424–429.

Hammer, R., Diesendruck, G., 2005. The role of dimensional distinctiveness in children and adults' artifact categorization. Psychol. Sci. 16 (2), 137–144.

Hammer, R., Hertz, T., Hochstein, S., Weinshall, D., 2007. Classification with positive and negative equivalence constraints: theory, computation and human experiments. In: Mele, F., Ramella, G., Santillo, S., Ventriglia, F. (Eds.), Brain, Vision, and Artificial Intelligence: Second International Symposium, BVAI 2007. Lecture Notes in Computer Science. Springer-Verlag Press, Berlin Heidelberg, pp. 264–276.

Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., Hochstein, S., 2008. Comparison processes in category learning: from theory to behavior. Brain Res. 1225, 102–118.

Hammer, R., Hertz, T., Hochstein, S., Weinshall, D., 2009a. Category learning from equivalence constraints. Cogn. Process. 10 (3), 211–232.

Hammer, R., Diesendruck, G., Weinshall, D., Hochstein, S., 2009b. The development of category learning strategies: what makes the difference? Cognition 112 (1), 105–119.

Hochstein, S., Ahissar, M., 2002. View from the top. Neuron 36 (5), 791–804.

Holyoak, K.J., 2008. Induction as model selection. Proc. Natl. Acad. Sci. 105, 10637–10638.

Krawczyl, D.C., Holyoak, K.J., Hummel, J.E., 2005. The one-to-one constraint in analogical mapping and inference. Cogn. Sci. 29, 797–806.

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 (6), 891–903.

Kéri, S., 2003. The cognitive neuroscience of category learning. Brain Res. Rev. 43 (1), 85–109.

Knowlton, B.J., Mangles, J.A., Squire, L.R., 1996. A neostriatal habit learning system in humans. Science 273, 1399–1402.

Kurmaran, D., Summerfield, J.J., Hassabls, Demis, Maguire, E.A., 2009. Tracking the emergence of conceptual knowledge during human decision making. Neuron 63, 889–901.

Maddox, W.T., Filoteo, J.V., 2001. Striatal contribution to category learning: quantitative modeling of simple linear and complex non-linear rule learning in patients with Parkinson's disease. J. Int. Neuropsychol. Soc. 7, 710–727.

Nomura, E.M., Maddox, W.T., Filoteo, J.V., Ing, A.D., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., Reber, P.J., 2007. Neural correlates of rule-based and information-integration visual category learning. Cereb. Cortex 17 (1), 37–43.

Reber, P.J., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2003. Dissociating explicit and implicit category knowledge with fMRI. J. Cogn. Neurosci. 15, 574–583.

Shohamy, D., Wagner, A.D., 2008. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron 60 (2), 378–389.

Shohamy, D., Myers, C.E., Onlaor, S., Gluck, M.A., 2004. The role of the basal ganglia in category learning: how do patients with Parkinson's disease learn? Behav. Neurosci. 118 (4), 676–686.

Sloutsky V.M., (in press). From perceptual categories to concepts: what develops? Cognitive Science.

Smith, E.E., Grossman, M., 2008. Multiple systems of category learning. Neurosci. Biobehav. Rev. 32, 249–264.

Spalding, T.L., Ross, B.H., 1994. Comparison-based learning: effects of comparing instances during category learning. J. Exp. Psychol. Learn. Mem. Cogn. 20 (6), 1251–1263.

Turk-Browne, N.B., Isola, P.J., Scholl, B.J., Treat, T.A., 2008. Multidimensional visual statistical learning. J. Exp. Psychol. Learn. Mem. Cogn. 34, 399–407.

Turk-Browne, N.B., Scholl, B.J., Chun, M.M., Johnson, M.K., 2009. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. J. Cogn. Neurosci.

Waldron, E.M., Ashby, F.G., 2001. The effects of concurrent task interference on category learning: evidence for multiple category learning systems. Psychon. Bull. Rev. 8, 168–176.

Zeithamova, D., Maddox, W.T., Schnyer, D.M., 2008. Dissociable prototype learning systems: evidence from brain imaging and behavior. J. Neurosci. 28 (49), 13194–13201.