Available online at www.sciencedirect.com
SCIENCE DIRECT
Brain and Cognition 60 (2006) 233–243
Brain and Cognition
www.elsevier.com/locate/b&c

Prototype-distortion category learning: A two-phase learning process across a distributed network

Deborah M. Little [a,*], Keith R. Thulborn [b]

[a] Center for Stroke Research, Departments of Neurology and Rehabilitation and Anatomy and Cell Biology, University of Illinois at Chicago, Chicago, IL 60612, USA
[b] Center for Magnetic Resonance Research, University of Illinois at Chicago, Chicago, IL 60612, USA

Accepted 30 June 2005
Available online 9 January 2006

Abstract

This paper reviews a body of work conducted in our laboratory that applies functional magnetic resonance imaging (fMRI) to better understand the biological response and change that occurs during prototype-distortion learning. We review results from two experiments (Little, Klein, Shobat, McClure, & Thulborn, 2004; Little & Thulborn, 2005) that provide support for increasing neuronal efficiency by way of a two-stage model that includes an initial period of recruitment of tissue across a distributed network that is followed by a period of increasing specialization with decreasing volume across the same network. Across the two studies, participants learned to classify patterns of random-dot distortions (Posner & Keele, 1968) into categories. At four points across this learning process subjects underwent examination by fMRI using a category-matching task. A large-scale network, altered across the protocol, was identified to include the frontal eye fields, both inferior and superior parietal lobules, and visual cortex. As behavioral performance increased, the volume of activation within these regions first increased and later in the protocol decreased. Based on our review of this work we propose that: (i) category learning is reflected as specialization of the same network initially implicated to complete the novel task, and (ii) this network encompasses regions not previously reported to be affected by prototype-distortion learning.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Category learning; Prototype-distortion; Functional magnetic resonance imaging

1. Introduction

Category learning has a large and rich history in cognitive science. However, with the increasing availability of neuroimaging there has been a recent increase in interest into the mechanisms involved in categorization and the theories that set out to characterize this skill. This paper reviews a body of work conducted in our laboratory that applies functional magnetic resonance imaging (fMRI) to better understand the biological response and change that occurs *during* the acquisition of one type of category learning: prototype-distortion learning. To do this, we will first review two neuronal responses that have been characterized as a result of learning and then define and review the mechanisms proposed to underlie prototype-distortion learning. We will then review findings from our laboratory, which present a description of the biological mechanisms that underlie prototype-distortion learning and relate them to this literature.

2. Changes in activation with learning

Petersen, Van Mier, Fiez, and Raichle (1998) have proposed two possible mechanisms, or an interaction between two mechanisms to explain the neuronal changes that are observed in response to learning.[1] This description of the

* Corresponding author. Fax: +1 312 355 5444.
E-mail address: little@uic.edu (D.M. Little).

[1] It is important to note that the focus of this review is on the learning of cognitive skills and not motor learning. This distinction is important because motor learning appears to have a distinct class of both behavioral and neuronal responses. A review of this literature is beyond the scope of this manuscript.

0278-2626/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandc.2005.06.004

changes that occur over the course of learning does not distinguish the specific networks involved in learning a specific task. The first possible mechanism proposed by Petersen and colleagues (1998) is that the neuronal network subserving behavioral improvements on a given task becomes more defined or efficient. With this increased efficiency, the associations between the regions of this network also strengthen (e.g. Toni, Rowe, Stephan, & Passingham, 2002). For the first mechanism, the establishment of more efficient connections would result in more specialized, or more well-defined, regions of brain activation. These regions of activation would implicate the same network whether early or late in the training program. Such neuronal changes, observed as reductions in the volume of activation as training progresses, have been reported for learning to associate an object with a location (Buchel, Coull, & Friston, 1999) and for artificial grammar learning (Thiel, Shanks, Henson, & Dolan, 2003). Although each of these reports shares the finding that the same network is implicated following practice, there is heterogeneity in the time course of these changes (Fletcher, Buchel, Josephs, Friston, & Dolan, 1999; Gonzalo, Shallice, & Dolan, 2000).

The second possible mechanism is that with successful learning a more task appropriate network is selected from a number of different networks (e.g. Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). This second mechanism would suggest that with practice a different set of cognitive operations processed over a different network is recruited to more efficiently accomplish the directed goal. The literature supporting the existence of both of these mechanisms is now presented briefly. The second mechanism, recruitment of a different network, might be realized in two ways. The first is that an entirely new network is utilized as learning progresses. The second outcome is that, rather than the creation of a new network, the newly acquired skill is incorporated as a component of an already existing expert skill and therefore mapped onto an already defined region known for its response to that expert skill. One example of this second outcome is that when subjects were trained to classify novel 3D characters by identifying specific features, similar to the requirements of face recognition, there was an increase in neuronal activity in fusiform gyrus, the same region that responded to faces (Gauthier et al., 1999). Additional support for this mechanism was observed when subjects were trained to read mirror-reversed text. As training on mirror-reading progressed, the same regions active in normal lexical processing and word recognition were activated (Poldrack, Desmond, Glover, & Gabrieli, 1998; Poldrack & Gabrieli, 2001).

This second mechanism is probably not distinct from the first but instead an additive factor. For example, during mirror reading reductions in some parts of the network are reduced whereas other areas show task-related increases. This reduction could either reflect specialization of the network as in the first mechanism or could reflect an independent process that no longer requires a large-scale neuronal network for task completion.

3. Category learning

Category learning, defined herein as the processes involved in extracting the rules that guide object classification, is rapid and can be carried out with apparent ease in healthy adults (Homa & Cultice, 1984; Palmeri & Flanery, 1999). Over the past decade advances in neuroimaging, in concert with well-characterized neuropsychological studies, have begun to allow more discrete characterization of the biological response to category learning. Two different classification schemes, one based upon the neuroimaging literature and one based upon the psychological literature, have been developed. The mechanistic classifications of the neuroimaging studies have allowed differentiation of the biological mechanisms of category learning that can be summarized into three distinct classes of biological response and based upon the regions implicated in completing the category leaning task: rule-based, information integration, and prototype-distortion (for a review, Ashby & Ell, 2001). Three slightly different behavioral models have also been proposed (Markman & Ross, 2003). These three classes of models differ from each other by the rules involved in creating a classification scheme. The three models are: rule-based, exemplar, and prototype models (Markman & Ross, 2003). The focus of both debates lies primarily in the mechanisms (both behavioral and biological) that are implicated in the process of extraction and definition of the critical features that define category membership.

Studies classified as rule-based tasks are those that can generally be completed successfully by application of a single rule that is based on a single stimulus dimension or a rule that can be easily explained. Neuroimaging data demonstrate a network of brain activation in regions involved in working memory (dorsolateral prefrontal cortex, DLPFC), hypothesis generation and testing (anterior cingulate), and for use of feedback (head of the caudate nucleus) (for a review see, Ashby & Maddox, 2005). Information integration studies are those that require subjects to learn associations between multiple stimulus dimensions (for example, Category A when the object is red but only on a blue background). The network involved during learning of information integration tasks also implicates regions involved in working memory (DLPFC), short-term memory (medial temporal lobes) as well as those regions implicated in processing feedback (tail of caudate nucleus) (Ashby & Maddox, 2005). Lastly, Ashby and Ell (2001) illustrate that those category learning tasks that utilize prototype-distortion generally show learning related changes in the visual cortices, suggesting that prototype-distortion learning is mediated by the same system involved in perceptual learning.

4. Definition of prototype-distortion learning

Prototype-distortion learning refers to a specific type of materials first introduced by Posner, Goldsmith, and Welton (1967). Specifically, subjects learn to back-classify

patterns of random dots created from distortions of prototype dot patterns. The categories are defined by a prototype with each category having a distinct prototype. The prototypes are created by a distribution of dots across a matrix (see Fig. 1). The dots in the prototype can either be intentionally placed so that an object or figure is loosely formed (for example, a square; Posner et al., 1967) or by a random distribution of dots within the matrix (Posner & Mitchell, 1967). Each prototype pattern is then "distorted" by shifting each dot within each prototype randomly in one of four directions (up, down, left, and right). This distortion can be re-applied to the prototype. Each time this distortion is applied the outcome is an exemplar for that prototype or category. Rate or speed of learning is directly related to the "degree" of distortion, or task difficulty (Posner et al., 1967).

The specific methods of the learning protocol are also important. Prototype-distortion tasks have been presented using one of two different methodologies. In the most popular version, subjects are given a study period in which they review distorted members of a single category (a). During the test period subjects are asked if a given stimulus is a member of that studied category or not a member of the studied category "a, not a." The second type of distortion task is when subjects are asked to differentiate between two or more categories "a,b." The distinction between these two tasks is important as they might reflect different skills. For the "a, not a" task subjects have to identify the common features of a single coherent category or have to learn only a single central feature. Importantly the exemplars that do not belong to that studied category are not from another single category but are created from numerous exemplars. For the "a,b" task learning is completed by not only learning the central component of each category and learning to distinguish between the categories.

Neuroimaging data further support the distinction between these two types of prototype-distortion tasks. For those studies employing the more popular "a, not a", learning related changes are largely localized to occipital cortex (Aizenstein et al., 2000; Reber, Stark, & Squire, 1998). In contrast, studies that have used "a,b" tasks found learning related changes not only in occipital cortex but also in prefrontal and parietal cortex (Seger et al., 2000). These findings lead to the suggestion that, although the stimuli are the same, the type of learning is different across these two methodologies. Specifically, when a single category is learned, more perceptual learning may be involved. When differentiation between categories is required more active hypothesis testing and spatial attention are recruited to complete the task and therefore is a more cognitive learning process.

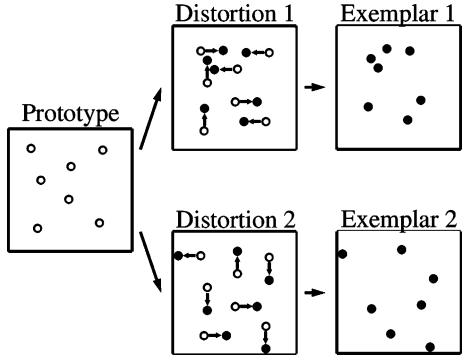

Fig. 1. Example prototype pattern of dots (prototype), the application of two distortions applied to the same prototype, and two exemplar patterns. *Note*. Although the dots and lines are shown here in black the actual stimuli used were inverted so that the dots and lines were white on a black background.

5. Mechanisms of prototype-distortion learning

Behaviorally, two competing theories of the cognitive operations that subserve prototype-distortion learning have arisen (for a review, Smith & Minda, 2001). To briefly summarize, one theory, initially proposed by Rosch (1975) and recently revised by Knowlton and Squire (Knowlton & Squire, 1993) suggests that participants create a running average of all patterns that fit within each prototype. This average prototype is modified whenever a new exemplar that fits the category is observed. The average is retained and used to judge incoming information. The specific exemplars used in training can be forgotten without producing a decline in performance (Squire & Knowlton, 1995). The second theory, the exemplar-based model, was originally proposed by Medin and colleagues (Medin & Schaffer, 1978; Medin & Smith, 1981) and has been recently revised by Nosofsky and Zaki (1998). The exemplar model proposed that traces of the actual exemplars used in training are stored in memory. Based upon this theory, judgments of incoming patterns are based upon the similarity of the to-be-categorized pattern with the stored exemplars.

Neuropsychological studies involving the prototype-distortion task have been limited to the "a, not a" methodology. Interestingly, this type of learning is intact in patients with amnesia (Knowlton & Squire, 1993); Alzheimer's disease (Sinha, 1999), and Parkinson's disease (Reber & Squire, 1999). These findings lend support to the conclusion that the type of learning that occurs on the "a, not a" tasks might rely more heavily on visual familiarity and perceptual learning and may not require active hypothesis testing (caudate, frontal regions) or medial temporal lobe structures.

The goal of the current paper is to integrate neuroimaging and behavioral results from our laboratory with the above literature that describes prototype-distortion learning. We set out to: (1) characterize the mechanism that describes the breadth of the network involved in this type of learning, and (2) characterize the change that occurs in this network with successful learning. It is this latter aim that distinguishes the current investigation from other recent investigations. To do this we utilized materials similar to Posner and colleagues (Posner & Keele, 1968; Posner & Mitchell, 1967) across two experiments and distributed the training and imaging over multiple days (imaging

across 4 days, training across the final 3 days). Because of the robust learning effects with these materials we choose to use a different task during the training and testing (imaging) sessions in an attempt to reduce or eliminate additional learning during the imaging sessions. The first imaging session in both experiments serves as a control to elucidate baseline activation patterns for completing the task. Prior to each of the remaining sessions subjects underwent a training session prior to imaging.

In the first experiment, we sought to interrogate the early learning process when the foundation for classifying the dot patterns is created. In the second experiment, we sought to characterize the entire learning process, from initial baseline performance to near perfect performance, which required additional practice, as compared to the first experiment, prior to each imaging session. Across both experiments, subjects initially imaged prior to any practice or training. Following this naïve state, participants were provided with practice prior to all remaining imaging sessions. The training involved performance of a 4-choice, speeded classification task with visual feedback following each trial. The testing during imaging was done with a 2-choice matching task without any feedback. The use of a different test and the lack of feedback during imaging reduced the likelihood of any additional change in behavior during imaging. Unlike previous investigations of prototype-distortion learning we controlled the amount and timing of training and the type and timing of feedback, characterized a baseline activation map prior to training, and sampled the biological response at multiple stages of learning. We additionally included a control task in the first experiment to characterize any changes in the biological response as a function of task and experimental familiarity including habituation to the scanning environment.

Our results demonstrate that under the current experimental conditions, prototype-distortion learning requires and affects activation across a large-scale biological network that includes frontal, parietal, and occipital regions. Further, by sampling across multiple days we demonstrate that prototype-distortion learning is reflected in two distinct biological phases that vary as a function of training and behavioral performance.

6. Materials and methods

The experimental paradigm and the procedures are described in detail in Little et al. (2004) and Little and Thulborn (2005). For both experiments, healthy subjects (Experiment 1: 17 right-handed young subjects (age range = 23–30 years, mean = 26.5 years); Experiment 2: 8 right-handed young subjects, (age range 24–31years; mean = 27.6 years)) participated after giving written informed consent. All had achieved at least 1 year of formal education beyond their undergraduate degrees (Experiment 1: range 17–20, mean = 18.2 years; Experiment 2: range = 17–19 years, mean = 17.3 years). All subjects were healthy and reported no history of neurological or psychiatric illness or injury. Each subject underwent 4 fMRI scanning sessions. Imaging on Day 1 followed a brief practice session without any training on the task. Imaging on Days 2, 3, and 4 followed a training session. Over the course of the protocol, subjects were trained to classify random patterns of dots into 4 categories.

6.1. Stimuli

The stimuli were modeled after Posner and Keele (1968) and consisted of "distorted" versions of "prototype" patterns of dots. Briefly, the "prototype" patterns are created by random distributions of 7 white dots across a grid (30×30 cells). Each of the prototype patterns was then "distorted" by means of moving each dot across the grid by 4 cells in either direction along the two orthogonal axes (up, down, left, and right). The direction of movement for each dot was random so that each time the distortion was applied, a new pattern of dots or "exemplar" was created (Fig. 1). For Experiment 1, 8 prototypes were created with 120 distorted versions or "exemplars" from each prototype. For Experiment 2, 4 prototypes were created with 80 exemplars from each prototype.

6.2. Training

The training consisted of a classification task which required subjects to determine to which of 4 categories a single pattern belonged (Experiment 1: 250 trials per day; 750 trials total; Experiment 2: 750 trials per day; 2150 trials total) (see Fig. 2). Dynamic visual feedback was presented following each trial and indicated accuracy and, if necessary, the correct category. Although subjects were instructed to determine category membership as quickly as possible, the total training was self-paced such that presentation of each trial was initiated by the subject. Response accuracy and latency (the time between onset of the exemplar and the key press) were recorded across both experiments.

6.3. Testing with fMRI

Two exemplars were projected side by side to the subject in the MRI scanner. Subjects were instructed to determine if the two exemplars were members of the same

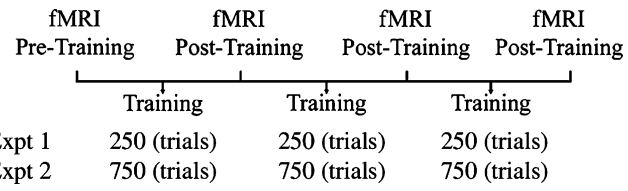

Fig. 2. Timeline of training and testing (imaging) for both Experiment 1 and 2. The primary difference between the experiments was the amount of training between imaging sessions with Experiment 1 having 250 trials of training per session and Experiment 2 having 750 trials of training.

category. Responses were recorded with a 2-choice button press (same/different). Each of the experimental paradigms was presented for 6.5 min and consisted of 0.5 min of the matching task interspersed with 0.5 min of central fixation. Although the paradigms were of identical design they differed in the class of exemplars presented. For Experiment 1, subjects were presented with a total of 3 experimental paradigms differing only in the type of exemplar presented: trained, untrained, and control. For Experiment 2 subjects were presented with only 2 experimental paradigms: trained and untrained. The *trained* paradigm involved presentation of exemplars that were also used during training. The *untrained* paradigm involved exemplars not used in training but that had been created from the same prototypes as used in the training sessions. The *control* paradigm used patterns novel to the subject that were created from prototypes not used in training. In these experiments, the untrained paradigm serves to control for explicit memory for materials and any effect of priming. The control paradigm serves to control for familiarity with the MRI environment.

6.4. Imaging parameters

A 3.0-T whole body scanner (Signa VHi, General Electric Medical Systems, Waukesha, WI) performing serial gradient echo, echo-planar imaging (epiRT, plane = axial, TR = 2999 ms, TE = 30.7 ms, flip angle = 90°, NEX = 1, Bandwidth = 62 kHz) was used for all image acquisition. Functional paradigms were then followed by acquisition of a high resolution 3D inversion recovery fast spoiled gradient recalled echo sequence (IRfSPGR, plane = axial, TR = 9 ms, TE = 2.0 ms, flip angle = 25°, NEX = 1, Bandwidth = 15.6 kHz, acquisition matrix = 256 × 256, FOV = 22 × 16.5 cm^2, slice thickness/gap = 1.5/0 mm/mm, slices = 124).

6.5. Imaging analysis

Detailed imaging analysis can be found in Little et al. (2004) and Little and Thulborn (2005). Briefly, data were screened for excessive head motion and excluded if total head motion across all conditions on any given day exceeded 3 mm (AFNI; Cox, 1996). All data were smoothed using a small isotropic Gaussian kernel (full width half maximum (FWHM) = 1.21). Image processing for each subject was carried out in FIASCO (Eddy, Fitzgerald, Genovese, Mockus, & Noll, 1996) and included the voxel-wise calculation of *t*-statistics for BOLD contrast between the category-matching condition and central fixation for each paradigm. AFNI (Cox, 1996) was then used for statistical thresholding, cluster thresholding, group averaging, normalization of the anatomy, and the region of interest (ROI) analysis. Separate group activation maps were calculated for each paradigm (Experiment 1: trained, untrained, control; Experiment 2: trained, untrained) for each day of testing. For the purpose of identifying significant clusters for further analyses, group images were calculated for each of these 12 activation maps for Experiment 1 and 8 activation maps for Experiment 2 for all voxels that exceeded the threshold. Cluster sizes exceeding 200 mm^3 or (8 voxels in original image space) were identified for further analyses (Forman et al., 1995). ROIs were defined based upon the Talairach and Tournoux stereotaxic atlas (Talairach & Tournoux, 1988) and anatomical boundaries consistent with the existing neuroimaging literature. Specifically, the ROIs were identified as part of the visuospatial networks and included left and right frontal eye fields (along the precentral sulci to include the adjacent gyri; Grosbras et al., 2001), the supplementary motor areas including the supplementary eye fields (anterior to the precentral sulcus and posterior to the caudate nucleus; Luna et al., 1998), left and right superior parietal lobule (superior and anterior to the intraparietal sulcus and posterior to the postcentral sulcus; Dassonville, Zhu, Ugurbil, Kim, & Ashe, 1997), left and right inferior parietal lobule (inferior and posterior to the intraparietal sulcus including the supramarginal gyrus; Frederikse, Lu, Aylward, Barta, & Pearlson, 1999), primary and secondary visual cortices (along the calcarine fissure to the cuneus and lingual gyrus; DeYoe et al., 1996), tertiary visual cortex (from the borders of V1/V2 to the middle occipital gyrus; DeYoe et al., 1996), and fusiform gyrus (from the mamillary body to the anterior tip of the parieto-occipital sulcus; Lee et al., 2002). All ROI analysis were conducted on the individual data.

7. Results

7.1. Behavioral findings: Training task

The behavioral results from the training are summarized in Fig. 3 for both Experiments 1 and 2. For both experiments there were significant effects of training such that there were overall increases in accuracy and overall decreases in response latency. Although the subjects for Experiments 1 and 2 were given differing amounts of training in a different time period, the results demonstrate similar findings. Specifically, there was no difference in the rate of learning, $p = .12$. There was a significant difference in response latency with participants in Experiment 1 taking significantly longer to classify patterns overall as compared with subjects in Experiment 2 who were given larger amounts of practice between sessions, $p = .043$.

7.2. Behavioral findings: Testing with fMRI

Data collected during the fMRI sessions followed the same trend as the training data collected prior to imaging. However, the behavioral data collected during fMRI scans allow investigation of baseline ability (prior to training), transfer effects (increase in accuracy for materials not used in training), and assimilation effects based on repeated

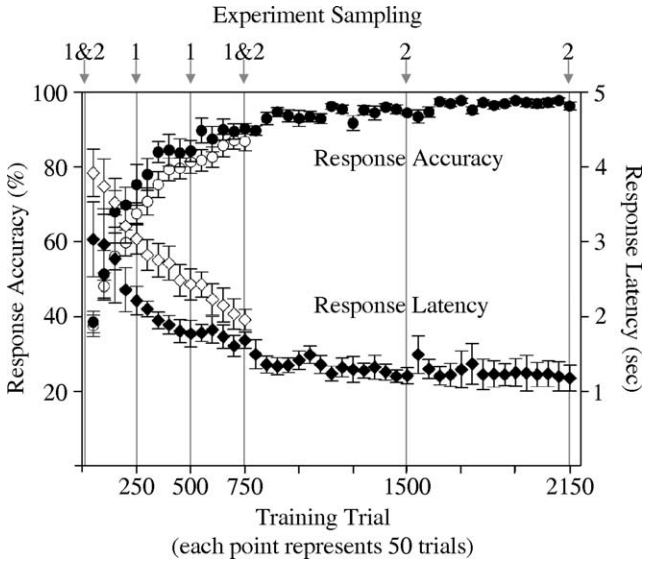

Fig. 3. Behavioral results from the training for Experiments 1 and 2 (indicated by notation above graph). Results are presented averaged for all 17 participants for accuracy (Experiment 1, open circles; Experiment 2, filled circles), and response latency (Experiment 1, open diamonds; Experiment 2, filled diamonds). Error bars represent 1 standard error. Data represented by filled symbols adapted from Little et al. (2004). Data represented by open symbols adapted from Little and Thulborn (2005).

exposure to the MRI (control, Experiment 1 only). Overall, subjects became more accurate as training progressed and that this training carried over to exemplars not used in training. However, there was no increase in accuracy for the control patterns (Experiment 1 only; presentation of novel exemplars from novel prototypes).

7.3. Regions of interest for fMRI analysis

The largest clusters of activation associated with the category-matching task were identified to include the supplementary eye fields, left and right frontal eye fields, left and right superior and inferior parietal lobules and left and right primary, secondary, and tertiary visual cortex. The center of each ROI for both Experiment 1 and 2 and their respective Broadmann areas are identified in Table 1. The patterns of activation for the category-matching task (for the trained exemplars) as compared to central fixation across the imaging sessions are presented in Fig. 4 (note that a group activation map is presented although all ROI analyses (Fig. 5) were completed on individual subjects).

7.4. Activation across learning

Significant overall effects across both Experiments as a function of day of training were observed for the frontal eye fields, superior and inferior parietal lobules and visual cortex (primary/secondary, and tertiary) as presented in Figs. 4 and 5. Significant effects in supplementary eye fields were only observed in Experiment 2 and will not be discussed further. Additional regions of activation were observed in certain individuals on various days of imaging (for example, dorsolateral prefrontal cortex and fusiform gyrus). However, we will focus on only those regions that were consistently active across all days of imaging. These regions are those with the largest and most consistent changes as a function of training and learning.

The change in activation across the protocol reflected two phases, with an initial increase in activation, termed

Table 1
Broadmann's areas (BA), Talairach coordinates, percent change in volume (%μL) and percent change in signal intensity (Experiment 1: activation differences after 750 trials compared to 250 trials; Experiment 2: activation differences after 2250 trials compared to 750 trials) for each ROI

Area	BA	Volume (μL)	Experiment 1					Experiment 2				
			Talairach			Change 750–250		Talairach			Change 2250–750	
			x	*y*	*z*	%μL	%sc	*x*	*y*	*z*	%μL	%sc
Frontal eye fields (FEF)	8											
Right		21089	−45	−5	34	5.4	>1	41.1	−6	43.9	37.3	>1
Left		22366	45	5	33	5.8	>1	−43.4	−4.7	42.7	38.7	>1
Visual cortex (V1 and V2)	17/18											
Right		22026	−9	80	−10	8.8	1.2	10.4	−88.3	−7.7	26.9	1.4
Left		22032	7	83	−9	7.2	>1	−11.4	−88.3	−7.7	27.2	1.2
Visual cortex (V3)	19											
Right		14663	−26	84	11	1.3	>1	23.9	−87.6	16.3	31.4	2.1
Left		14662	27	83	16	2.6	>1	−24.9	−87.6	16.7	33.8	1.8
Supplementary motor (SMA and SEF)	6											
Right		4085	−3	−3	55	3.2	>1	5.3	8.1	52.8	18.7	>1
Left		4218	3	−6	53	4.1	>1	−4.8	8.2	52.9	23.1	1.2
Inferior parietal lobule (IPL)	40											
Right		20276	−39	40	39	1.3	>1	−48.4	−40.7	38.7	47.7	>1
Left		20291	50	36	46	2.4	>1	−48.5	−40.7	38.7	51.4	>1
Superior parietal lobule (SPL)	7											
Right		5907	−31	65	47	2.3	>1	27.2	−59.4	52.7	35.4	1.1
Left		6002	43	60	51	1.4	>1	−27.2	−59.4	52.7	46.5	>1

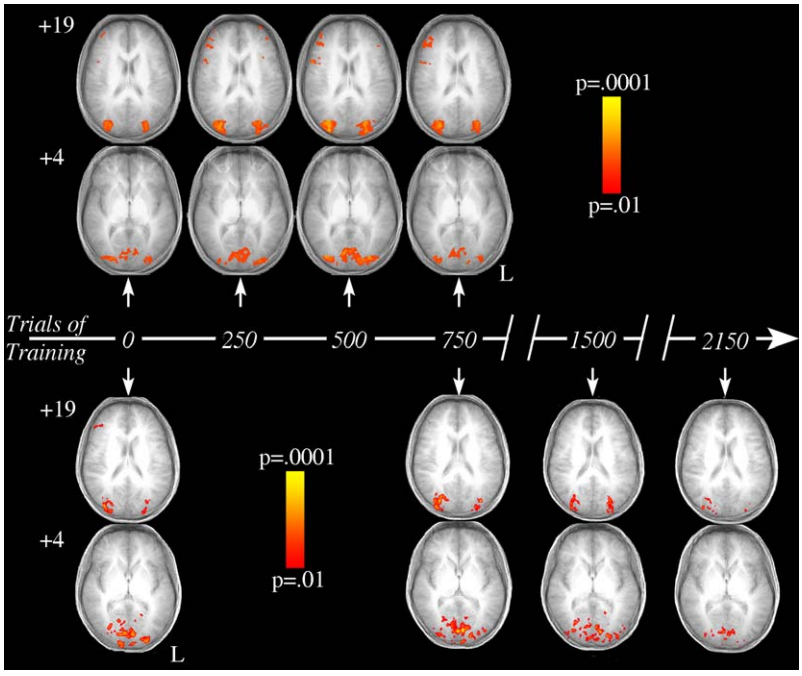

Fig. 4. Representative axial slices of the whole brain activation maps associated with the category-matching task as compared to central fixation for the each day of imaging for Experiments 1 (top) and 2 (bottom). Activation maps are aligned horizontally when imaging was completed at the same relative time (with regard to amount of training prior to imaging). Images are presented according to radiological convention (right side of image presented on the left).

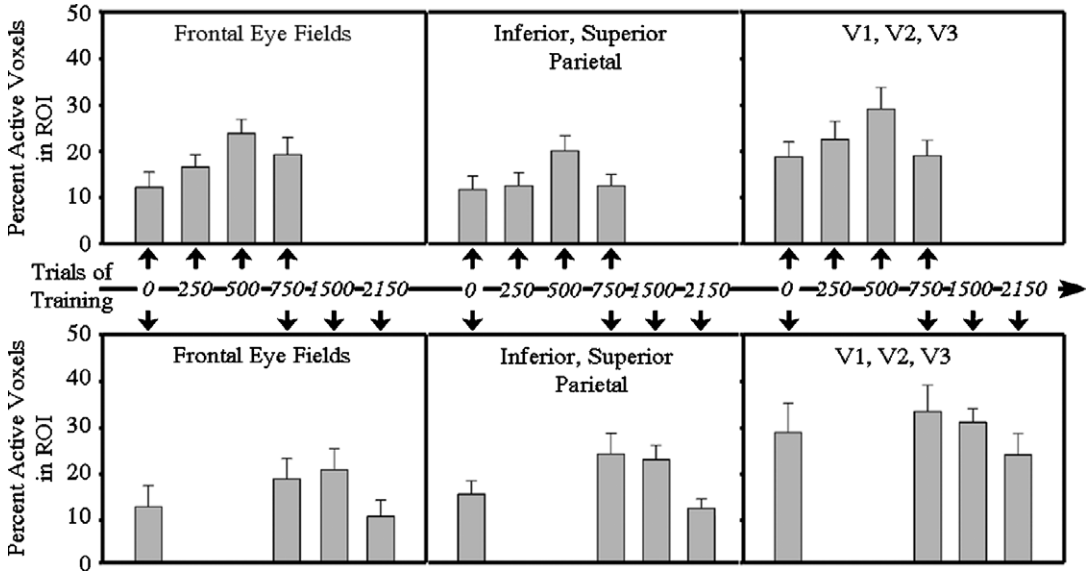

Fig. 5. Volumes of activation (percent of active voxels as a function of the volume of entire ROI) are presented for the six regions of interest that demonstrated change as a function of the training and were consistently active across subjects for Experiment 1 (top row) and Experiment 2 (bottom row). The bars in each graph are aligned between the two experiments to represent MRI sampling relative to training.

recruitment, and a later decrease in activation, termed specialization. The increase in activation, or recruitment period, was prolonged by smaller amounts of practice (i.e., specialization was delayed) in Experiment 1 as compared to Experiment 2. This observation, that differences in methodology between the two experiments affected the time course of change in activation, was significant for the frontal eye fields, $F(7,100) = 2.81$, $p = .04$, secondary and tertiary visual cortices, $F(7,100) = 4.99$, $p = .003$, and for both the inferior, $F(7,100) = 8.08$, $p < .001$ and superior parietal lobules $F(7,100) = 9.6$, $p < .001$. Although there were significant effects of learning observed in the primary visual cortex, these effects were consistent across both experiments and therefore the differences in methodology did not change the patterns of activation in primary visual cortex. However, these significant differences between Experiment 1 and 2 are removed if comparisons are made with regard to practice rather than imaging session. For example, there is no differ-

ence in the time course of specialization for the frontal eye fields if the third imaging session in Experiment 1 is compared with the second imaging session in Experiment 2, each occurring following 750 trials of practice.

It is important to note that although there were significant decreases in the overall volume of activation within the ROIs, there were no significant changes in either the distribution (Experiment 1: $\chi^2 > .09$; Experiment 2: $\chi^2 > .06$) or the magnitude of the BOLD signal change as a function of training (Experiment 1: F tests with $p > .11$; Experiment 2: F tests with $p > .08$).

8. General discussion

The current paper summarizes two complementary data sets that, when taken together, provide two main observations. In Experiment 1, there was an increase in activation from Day 1 to Day 2 and continued through Day 3. This increase was followed by a decrease in activation on the final day of imaging for the trained paradigms. There was no decrease in activation observed for either of the untrained or control paradigms. For Experiment 2, there was an increase in activation from Day 1 to the testing session immediately following the first training session (Day 2). This was followed by a decrease in volume of activation as training progressed for the learned exemplars in all ROIs as training progressed to Day 3 and Day 4.

The results demonstrate a network of activation that is more extensive than previously reported for prototype-distortion learning (Reber et al., 1998; Vogels, Sary, Dupont, & Orban, 2002). Our data demonstrate a network of activation that is affected by training that includes regions involved in visual processing, object recognition, spatial attention, and the control of eye movements. The learning related effects were observed as initial increases in activation which paralleled the largest increases in behavioral accuracy immediately following the initial training session. With further training (after 750 trials in Experiment 1 and 1500 in Experiment 2) activation in these regions was significantly reduced.

Similarly, Reber and colleagues (1998) report a large-scale network of activation during a prototype-distortion task that included frontal, parietal, and visual cortices. However, unlike the current studies, the only learning related changes were present in the visual cortices. Ashby and Ell (2001) interpret the learning related reduction in visual cortex activation and the absence of learning-related changes in activation across the frontal and parietal regions to indicate that the prototype-distortion task reflects perceptual learning rather than rule-based or information-integration learning. Both of these mechanisms rely on hypothesis testing and explicit memory. In contrast reports by Vogels et al. (2002) and Aizenstein et al. (2000) demonstrated learning related changes using the prototype-distortion task that were not limited to posterior occipital regions and included frontal and parietal regions of activation. Activation in visual areas is common across all studies.

The question is then as to what factors are necessary during learning to engage the frontal and parietal regions of the network. The implication of frontal activation has been interpreted to represent explicit memory processes (Aizenstein et al., 2000; Reber et al., 1998). In the case of Aizenstein et al., pre-frontal activation (DLPFC) was only reported when subjects were explicitly categorizing dots into categories. No frontal activation was observed when the control task (color judgments of the same patterns) was carried out. However, in the current studies, subjects were always explicitly categorizing dot patterns yet DLPFC was not affected by training and was not consistently active across subjects. From the current results, activation in DLPFC is not directly linked to explicit processes nor is it affected by our training protocol. We believe the discrepancy in DPLFC activation across the reviewed studies and our own results is more tightly linked to explicit hypothesis testing. In the current studies, training and feedback was only experienced outside of the scanner. The task used in the scanner was designed to test category knowledge rather than allow continued learning during the fMRI sessions. This design allowed us to characterize the biological response at multiple times during learning without changing the learning process per se. The question of whether activation in DLPFC is feedback dependent and mediated by the quality of feedback cannot be addressed by our data. However, the data across all studies suggests that prototype-distortion learning involves more than perceptual learning and implicates more regions than just visual cortex. The observed network may be better classified and modeled with information-integration models of category learning rather than those describing perceptual learning. Experimental protocols, such as those of Aizenstein and Vogels, require discrimination between multiple categories. It could be the case that when subjects are required to study and discriminate patterns only belonging to a single category, perceptual learning is implicated. In this case, recognition and discrimination of patterns may be related to familiarity rather than higher order learning per se. Furthermore, the current studies demonstrate that category learning cannot be characterized by a single activation map. The biological basis of category learning and specifically, prototype-distortion learning is temporally sensitive to the stage of learning.

If prototype-distortion learning engages mechanisms that include but are not entirely explained by perceptual learning task then the question arises as to what processes are involved during distortion learning. The prototype based theory of Knowlton and Squire (1993) would create the following biological scenario: If each exemplar within a given category was used to create a running average, then the processing load would be heavy initially when the average pattern could change significantly with the addition of each new exemplar. As an increasing number of exemplars or samples were added to this average, the prototype would change very little, producing a situation of reduced processing demands. The exemplar based theory of Nosofsky and Zaki (1998) would also predict heavy initial processing for

storage and judgments of similarity based upon few exemplars. As the knowledge base was expanded, the chances of great similarity between any new exemplar with one already stored would increase and processing would be reduced. Further differentiation of these models would require an imaging protocol that varied not only the presence or absence of exemplars in training but also the degree of distortion for the exemplars. The results presented fit both proposed modes of prototype-distortion learning but which can now be refined to include an underlying two-stage distributed process that initially demonstrates recruitment of tissue followed by specialization.

We interpret the initial increases in activation to reflect recruitment during the initial development of prototype or category knowledge. Once this knowledge has been established, a decrease or specialization, of activation across the network is observed. These findings support the conclusion that the type of learning that occurs with the development of the categorization skill is one of specialization as described by Petersen and colleagues (1998).

It is important to note that although the materials are similar to those previously used to characterize prototype-distortion learning, our methods are quite different. First, studies of prototype-distortion have generally made use of either one or two categories. In these studies subjects were required to either simply determine category membership (Category X, Other) or distinguish between two categories (X,Z). For both Experiments summarized above (for details, Little et al., 2004; Little & Thulborn, 2005) subjects were required to distinguish between four separate categories (A,B,C,D). Second, the difficulty of the categorization task was such that learning was accomplished over a relatively large number of trials (Experiment 1, 750 trials; Experiment 2: 2150 trials) across multiple days. Most other protocols involving prototype-distortion utilize a single MRI session preceded by a single study session. Our studies allowed a finer gradation of measurements across the learning process. Third, we characterized the biological response separately to items used in training, items created with the same rules but not used in training, and a set of control materials to characterize changes in baseline over a 4-day protocol.

One possible confound in interpreting the effects of learning on the biological response is that of changes in the cognitive duty cycle (Poldrack, 2000). The cognitive duty cycle is the time or amount of biological resources required to complete a cognitive task. In the case of the current studies, as the time to classify an item decreased across the protocol the corresponding duty cycle was also reduced. A reduction in duty cycle across the protocol would be observed as a reduction in the average signal intensity across the fMRI task. In fact, although the time to complete the task decreased (as indicated by reduced reaction time measurements), the signal change remained constant across training. Additional support against an interpretation of cognitive duty cycle dominating the BOLD response and being reflected in volumetric changes comes from the observation that the largest reduction in response latency occurs immediately after the first training session when the volume of activation is increased.

Following this methodological point, it is important to acknowledge the limitations of selection criteria for the current analysis. In both experiments, the SC (signal change) was calculated in two ways. In the first the total signal was calculated across the entire ROI which includes voxels that did not exceed our threshold which we applied to the volume calculations. The result of this analysis paralleled the volumetric findings; as volume increased the SC increased, as volume decreased SC decreased. However, to address the question of changes within those supra-threshold voxels that survived the volume analysis we also calculated SC only within these supra-threshold voxels. It is this latter analysis that would address questions as to whether the voxels that remain active are "specialized" for the experimental stimuli. This analysis produced the result that BOLD signal did not change across the protocol while the volume of activated voxels did change. From this finding we can conclude that the change in volume of activation cannot be a change in the magnitude of the BOLD signal but instead must reflect a change in the extent of activation across the brain (i.e., volume of activation reflects the volume of tissue involvement during the cognitive process). This observation also has one additional implication on the theory that prototype-distortion learning reflects a perceptual process (Ashby & Ell, 2001).

Perceptual learning, as with any type of learning, refers to changes in performance that occur as a function of practice or experience with a task or with specific materials. Perceptual learning can be distinguished from other types of learning because it is believed to alter receptive field mapping in primary sensory cortical regions (Dosher & Lu, 1999). In the case of prototype-distortion learning, the theory of Ashby and Ell (2001) proposes that human subjects learn to associate a category by the distribution of dots in space. With experience, receptive fields in visual cortex become tuned to clusters of dots and show a greater response over learning (Dosher & Lu, 1999). If this tuning did occur for prototype-distortion learning one would hypothesize an increase in specific regions of visual cortex through training. In the current two experiments not only did the signal intensity within those supra-threshold voxels not change but a voxel-by-voxel analysis indicated that not even a single voxel showed an increase in signal intensity over the course of the protocol. We believe this finding argues against the perceptual learning hypothesis with regard to prototype-distortion learning in the current experimental protocol. It is also important to acknowledge that the current study has a focus on visual learning. It could easily be imagined (as with perceptual learning) that if the stimuli were of an auditory nature, the primary changes would occur in auditory cortex and implicate a different network. Understanding the changes that occur during learning, not just identification of the areas involved in learning, is important to any biological model of learning.

Our data lend support to the model that successful category learning and the concomitant development of increasing efficiency appears to consist of two distinct yet related stages. This two-stage process is initially composed of tissue recruitment that is then followed by specialization. The activation maps for prototype-distortion learning implicates a large-scale network that cannot be explained by perceptual learning. This finding is significant because prototype-distortion learning appears intact across a wide variety of patient populations that show decrements in performance on other category learning tasks (for a review, Ashby & Ell, 2001). This leads to a series of questions about the effects of disease and what components of the network are spared to allow intact learning in the presence of other behavioral decrements. Answers to such questions may provide insight into the plasticity of the networks that compensate for disease processes.

Acknowledgments

This work was supported by NIH grant PO1 NS 35949 from the National Institute of Neurological Disorder and Stroke and a grant from the Alzheimer's Association ZEN-99-1790. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of either the National Institute of Neurological Disorder and Stroke or the Alzheimer's Association.

References

Aizenstein, H. J., MacDonald, A. W., Stenger, V. A., Nebes, R. D., Larson, J. K., Ursu, S., et al. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience, 12*, 977–987.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences, 5*, 204–210.

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

Buchel, C., Coull, J. T., & Friston, J. T. (1999). The predictive value of changes in effective connectivity for human learning. *Science, 283*, 1538–1541.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*, 162–173.

Dassonville, P., Zhu, X. H., Ugurbil, K., Kim, S. G., & Ashe, J. (1997). Functional activation in motor cortex reflects the direction and degree of handedness. *Proceedings of the National Academy of Sciences of the United States of America, 94*, 14015–14018.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wiser, J., Cox, R., et al. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America, 93*, 2382–2386.

Dosher, B., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research, 39*, 3197–3221.

Eddy, W. F., Fitzgerald, M., Genovese, C. R., Mockus, A., & Noll, D. C. (1996). Functional image analysis software—Computational olio. In A. Prat (Ed.), *Proceedings in computational statistics*. Heidelberg: Physica-Verlag.

Fletcher, P., Buchel, C., Josephs, O., Friston, K., & Dolan, R. (1999). Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex, 9*, 169–178.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33*, 636–647.

Frederikse, M. E., Lu, A., Aylward, E., Barta, P., & Pearlson, G. (1999). Sex differences in the inferior parietal lobule. *Cerebral Cortex, 9*, 896–901.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, P. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*, 568–573.

Gonzalo, D., Shallice, T., & Dolan, R. (2000). Time-dependent changes in learning audiovisual associations: A single-trial fMRI study. *NeuroImage, 11*, 243–255.

Grosbras, M. H., Leonards, U., Lobel, E., Poline, J. B., LeBihan, D., & Berthoz, A. (2001). Human cortical networks for new and familiar sequences of saccades. *Cerebral Cortex, 11*, 936–945.

Homa, D., & Cultice, J. C. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology—Learning, Memory and Cognition, 10*, 83–94.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262*, 1747–1749.

Lee, C. U., Shenton, M. E., Salisbury, D. F., Kasai, K., Onitsuka, T., Dickey, C. C., et al. (2002). Fusiform gyrus volume reduction in first-episode schizophrenia: a magnetic resonance imaging study. *Archives of General Psychiatry, 59*, 775–781.

Little, D. M., Klein, R., Shobat, D. M., McClure, E. D., & Thulborn, K. R. (2004). Changing patterns of brain activation during category learning revealed by functional MRI. *Cognitive Brain Research, 22*, 84–93.

Little, D. M., & Thulborn, K. R. (2005). Correlations of cortical activation and behavior during the application of newly learned categories. *Cognitive Brain Research, 25*, 33–47.

Luna, B., Thulborn, K. R., Strojwas, M. H., McCurtain, B. J., Berman, R. A., Genovese, C. R., et al. (1998). Dorsal cortical regions subserving visually guided saccades in humans: An fMRI study. *Cerebral Cortex, 8*, 40–47.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592–613.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology—Human Learning and Memory, 7*, 241–253.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition memory in amnesic and normal individual: An exemplar-based interpretation. *Psychological Science, 9*, 247–255.

Palmeri, T. J., & Flanery, M. A. (1999). Learning about categories in the absence of training. Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science, 10*, 526–530.

Petersen, S. E., Van Mier, H., Fiez, J., & Raichle, M. (1998). The effects of practice on the functional anatomy of task performance. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 853–860.

Poldrack, R. A. (2000). Imaging brain plasticity: Conceptual and methodological issues—A theoretical review. *NeuroImage, 12*, 1–13.

Poldrack, R. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). The neural basis of visual skill learning: An fMRI study of mirror reading. *Cerebral Cortex, 8*, 1047–1056.

Poldrack, R. A., & Gabrieli, J. D. E. (2001). Characterizing the neural mechanisms of skill learning and repetition priming: Evidence from mirror reading. *Brain, 124*, 67–82.

Posner, M. I., Goldsmith, R., & Welton, K. E. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology, 73*, 28–38.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353–363.

Posner, M. I., & Mitchell, RF. (1967). Chronometric analysis of classification. *Psychological Review, 74*(5), 392–409.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of*

the National Academy of Sciences of the United States of America, 95, 747–750.

Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioral Neuroscience, 113*, 235–242.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7*, 192–238.

Seger, C. A., Poldrack, R. A., Prabhakaran, V., Zhao, M., Glover, G. H., & Gabrieli, J. D. E. (2000). Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia, 38*, 1316–1324.

Sinha, R. R. (1999). Neuropsychological substrates of category learning. *Dissertation Abstracts International, Section B: Science and Engineering*, *60*(5B), 2381 (UMI: AEH9932480).

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology—Learning, Memory and Cognition, 27*, 984–1002.

Squire, L. R., & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences of the United States of America, 93*, 13515–13522.

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

Toni, J., Rowe, J., Stephan, K. E., & Passingham, R. E. (2002). Changes of cortico-striatal effective connectivity during visuomotor learning. *Cerebral Cortex, 10*, 1040–1047.

Thiel, C. M., Shanks, D. R., Henson, R. N., & Dolan, R. J. (2003). Neuronal correlates of familiarity-driven decisions in artificial grammar learning. *Neuroreport, 14*, 131–136.

Vogels, R., Sary, G., Dupont, P., & Orban, G. A. (2002). Human brain regions involved in visual categorization. *NeuroImage, 16*, 401–404.