Functional MR imaging of semantic information processing and learning-related effects using psychometrically controlled stimulation paradigms

Manfred Spitzer \(^a\), Matthias E. Bellemann \(^b\), Thomas Kammer \(^a\), Friedemann Gückel \(^c\), Udo Kischka \(^d,e\), Sabine Maier \(^a\), Andreas Schwartz \(^d\), Gunnar Brix \(^b\)

\(^a\) Sektion Experimentelle Psychopathologie, Psychiatrische Universitätsklinik, Heidelberg, Germany
\(^b\) Forschungsschwerpunkt Radiologische Diagnostik und Therapie, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany
\(^c\) Institut für Klinische Radiologie, Klinikum Mannheim, Klinische Fakultät II der Universität Heidelberg, Heidelberg, Germany
\(^d\) Neurologische Klinik, Klinikum Mannheim, Klinische Fakultät II der Universität Heidelberg, Heidelberg, Germany
\(^e\) Neurologische Universitätsklinik, Magdeburg, Germany

Accepted 12 March 1996

Abstract

Functional magnetic resonance imaging (fMRI), in conjunction with carefully designed, psychometrically optimized stimulation procedures, was used to investigate the relation between brain activation and the processing of word associations. A semantic discrimination task of word-pair similarity was performed by normal subjects \((n = 17)\) within a clinical 1.5-Tesla whole-body MRI system. A color similarity task of psychometrically equivalent difficulty, as indicated by behavioral data acquired online during fMRI, served as active control condition. Comparisons between tasks dramatically improved results compared to comparisons between task and resting condition. The language paradigm selectively activated left frontal and left fronto-temporal areas. Cortical activation during the semantic task decreased significantly over three runs of the same word list and was paralleled by decreased reaction times. No such changes were observed in the active control condition indicating selective learning of the language task only. When combined with psychological activation schemes and the acquisition of behavioral data, fMRI represents a powerful tool for the study of brain–behavior interaction.

Keywords: Cognitive neuroscience; Functional magnetic resonance imaging (fMRI); Psycholinguistics; Brain activation; Learning

1. Introduction

The living brain is never at rest. This bears one of the most pressing problems for any attempt to image brain function, i.e. what state to use as reference against which activation can be measured. Whereas for the study of simple sensory or motor phenomena subjects can be instructed to close their eyes or not to move, the problem is far from trivial when higher cognitive functions are under scrutiny. For example, when subjects are instructed to close their eyes, this may facilitate the occurrence of internally generated mental images of colors and shapes or even of entire scenes. According to recent research using positron-emission tomography (PET), this activity under conditions of imagery is produced by cortical mechanisms very similar to those that produce visual images [23,24]. By analogy, the investigation of attentional, language-related or thought processes must face the problem that awake subjects will always think (internally speak of), and attend to, something. Hence, attempts to find cortical areas activated by thinking, speaking or attending have to face the problem of the adequate ‘resting’, or control, condition. This problem has been addressed in psychological experiments by the use of the difference method with the following rationale: measure a simple process; then measure another process related to the first one but different in one new aspect. The difference of the two measurements is due to this new aspect, which thereby can be quantified [40]. Notice that the difference method was designed to investigate differences between two (slightly different) tasks, but not between a ‘resting state’ and a task.
In their elegant study of language processes, Petersen and coworkers [35] have demonstrated the applicability of the difference method to PET. This approach can be adapted to functional magnetic resonance imaging (fMRI), an important new tool for the study of the working brain. Whereas the studies of patients with lesions suggest that the brain is organized into distinct areas of relative functional autonomy and specialization, such studies do not allow to observe these areas in action. Only functional imaging methods reveal how complex cognitive tasks are performed by the intricate interplay of a number of different brain areas, each carrying out a specific computational aspect of the general task. Functions, such as language, attention, thought or vision, are no longer said to be carried out by one or two ‘centers’, but rather by ‘networks’ comprising different cortical and subcortical areas [12,30]. Until a few years ago, investigations of the structure-function relationship of these areas involved the use of radioactive materials and had a limited spatial (5 to 10 mm) and low temporal (30 s) resolution. In contrast, fMRI is noninvasive, and – in principle – a better spatial and temporal resolution than PET. Like PET, fMRI provides only an indirect measure of neuronal activity, as it detects changes in blood flow, using deoxygenated blood as an endogenous contrast agent.

Functional MRI has been successfully used to demonstrate cortical activation by visual stimulation and motor tasks [3,6–9,14–16,18,22,25,28,31,45–47,55–57], and a number of studies have gone beyond sensorimotor phenomena into the realm of higher cognitive functioning, especially language (e.g. [4,5,13,21,27,43,48,58]). Because of the above mentioned advantages of fMRI and the considerably large base of installed clinical MR machines, and because of the growing interest in cognitive neuroscience [19], it can be expected that this methodology is used by an increasing number of research groups for the study of higher cognitive functions. When this is done, behavioral and image data should be recorded at the same time to make the interpretation of functional image data as simple and straightforward as possible. Hence, task construction as well as behavioral data acquisition become crucial [59].

Almost any task can be constructed in such a way that the subject has to make a response. A number of reasons suggests that whenever possible, response measures should be taken. First of all, neural activity, as indicated by functional activation studies, by itself has no unambiguous interpretation. While neurons may signal activation, they may also modulate the activity of other neurons or they may even signal inhibition [39]. Moreover, the relation between the fMRI signal and neuronal firing is not completely understood. In order to assess the significance of imaged activity for a specified mental operation, a response measure that taps this operation should be taken, and it should be used to discuss imaging results. Second, response measures are almost indispensible when patient populations rather than highly motivated human control subjects are studied in order to make sure that the task in question is performed properly at all.

The most widely employed response measure in psychology is reaction time (RT). This general indicator of the speed of mental operations [38] can be readily measured during scanning procedures, if the task is set up such that it requires a response. The subject may be asked to merely detect the presence of a stimulus (simple RT paradigm) or to make a decision (forced choice RT paradigm). Sometimes, the subject may be asked to respond only to some stimuli, i.e. to make a decision and to indicate it by either responding or not responding (go/no-go paradigm). In addition to RTs, error rates (ERs) may be monitored, as some subjects may respond fast at the expense of making more errors than others who may take more time but also more care (speed-accuracy tradeoff). In the context of the present study, it should be added that when subjects learn a task, they tend to do it faster and more accurate. This renders RTs and ERs as useful measures of the amount of learning taking place during the prolonged and/or repeated performance of a given task.

In this paper, we try to exemplify the procedures that should make the study of higher cognitive functions possible at many clinical MRI sites. We argue that stimulation and data acquisition procedures have to be carefully designed with regard to the psychometric aspects of the tasks. Small-scale experiments for the validation of stimulation and data acquisition procedures in general can be performed off-line in the psychological laboratory. Most importantly, we demonstrate the benefits of using appropriate control tasks and of implementing adequate behavioral response measures. ‘Task vs. no-task’ designs should not be used. Finally, we provide evidence for the usefulness of time-course data for the study of distributed cortical networks involved in the processing of language.

2. Materials and methods

2.1. General setup

The setup allowed computerized stimulus presentation and response measurement similar to standard tachistoscopic experiments used in experimental psychology. All studies were carried out on a clinical MR system (see below) with additional hard- and software for stimulus presentation and response measurement purposes. The presentation of stimuli as well as the collection of response data was controlled by an Apple Macintosh microcomputer, located outside the scanning room, running special purpose software (MacLab 2.0; cf. [11]) customized for each experiment. Stimuli were generated with standard image processing software (Adobe Photoshop 2.5; Adobe Systems, Mountain View, CA). A videoprojector (Sharp Vision XG 3800E) was used to project the stimuli into the
scanning room on a translucent screen which could be viewed by the subjects from within the scanner using prism spectacles. Stimuli were set up vertically and horizontally flipped to compensate for back-projection and prism-use. This setup allowed for viewing of stimuli obtaining a visual angle of \( \approx 24^\circ \) in the horizontal and \( 12^\circ \) in the vertical direction from within the scanner.

The measurement of RTs implies a motor response by the subject, which inevitably is accompanied by somatosensory impressions. In other words, any motor response causes cortical and/or subcortical activity which may interfere with the activation supposedly due to the task. Hence, care was taken to utilize the appropriate effector body part for making the response. In order to minimize interference of the sensorimotor component of the response with imaging of language-related cortical areas, the left foot was chosen to make responses, as its cortical motor and sensory representation is located relatively far from the language-related areas of normal right-handed volunteer subjects.

From a technical point of view, RTs and ERs may best be measured using a fiberoptic system that does not interfere with the MRI scanner. However, we tested the feasibility of a wooden/plastic two-way switch hooked to a computer via phone wire, and were unable to detect any electromagnetic interference with the MR imaging technology. The subject could make ‘yes’ or ‘no’ decisions by turning his or her left foot either to the right or to the left. This was equivalent to hitting one of two keys of the keyboard – a signal utilized by the computer to obtain RTs and ERs online.

### 2.2. Subjects

All subjects \((n = 17)\) were recruited from the Heidelberg area. They had no history of neurological or psychiatric disorder and gave informed consent to the MR examination prior to the investigation, after the aims of the study and the experimental procedures had been fully explained to them. Handedness was scored using the Edinburgh Inventory [34]. Demographic and handedness data of subjects, together with results of the behavioral measurements, are presented in Table 1.

### 2.3. Imaging protocol

A conventional gradient system equipped 1.5-Tesla whole-body MRI system (MAGNETOM 63 SP, Siemens, Erlangen, Germany) was used for image generation. For imaging of cortical regions, we used a circular-polarized head coil for radiofrequency transmission and detection at 63.64 MHz. Head fixation was accomplished by bitemporal cushions and by an additional head fixation device (bitebar). To localize the planes of interest, we first acquired 19 sagittal \( T_1 \)-weighted spin-echo images using the following data acquisition parameter settings: repetition time, \( T_R = 600 \) ms; echo time, \( T_E = 15 \) ms; matrix size, \( MA = 256 \times 128 \); field-of-view, \( FOV = 300 \) mm; slice thickness, \( TH = 5 \) mm; number of excitations, \( NEX = 1 \).

Based on these images, three (Expt. 1) or four (Expt. 2) adjacent planes (\( TH = 5 \) mm; slice distance, \( SD = 0.5 \) mm; \( FOV = 220 \) mm) were placed such that the cortical areas under investigation could be expected to be maxi-

---

**Table 1**

Demographic, behavioral and imaging data for Expts. 1 and 2

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age, gender</th>
<th>Handedness score</th>
<th>Number of slices/rest period</th>
<th>RT semantic (mean ± S.D.) (ms)</th>
<th>RT color (mean ± S.D.) (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expt. 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M.P.</td>
<td>24, m</td>
<td>+83</td>
<td>3/yes</td>
<td>1726 ± 692</td>
<td>1448 ± 579</td>
</tr>
<tr>
<td>J.R.</td>
<td>27, f</td>
<td>+50</td>
<td>3/yes</td>
<td>2215 ± 720</td>
<td>1739 ± 573</td>
</tr>
<tr>
<td>K.S.</td>
<td>24, m</td>
<td>+100</td>
<td>3/yes</td>
<td>1751 ± 668</td>
<td>1449 ± 471</td>
</tr>
<tr>
<td>A.L.</td>
<td>25, m</td>
<td>+83</td>
<td>3/yes</td>
<td>2168 ± 979</td>
<td>1945 ± 880</td>
</tr>
<tr>
<td>D.S.</td>
<td>26, m</td>
<td>+69</td>
<td>3/yes</td>
<td>1634 ± 896</td>
<td>2916 ± 1335</td>
</tr>
<tr>
<td>M.O.</td>
<td>39, m</td>
<td>+100</td>
<td>3/yes</td>
<td>1076 ± 293</td>
<td>1301 ± 376</td>
</tr>
<tr>
<td>L.A</td>
<td>27, m</td>
<td>+57</td>
<td>3/yes</td>
<td>2907 ± 1148</td>
<td>3071 ± 1228</td>
</tr>
<tr>
<td><strong>Expt. 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.M.</td>
<td>26, f</td>
<td>+100</td>
<td>4/no</td>
<td>2388 ± 896</td>
<td>2717 ± 823</td>
</tr>
<tr>
<td>U.R.</td>
<td>33, m</td>
<td>+100</td>
<td>4/no</td>
<td>1677 ± 657</td>
<td>1861 ± 681</td>
</tr>
<tr>
<td>R.R.</td>
<td>28, m</td>
<td>+83</td>
<td>4/no</td>
<td>2522 ± 480</td>
<td>2262 ± 385</td>
</tr>
<tr>
<td>A.H.</td>
<td>31, f</td>
<td>+100</td>
<td>4/no</td>
<td>2073 ± 821</td>
<td>2024 ± 713</td>
</tr>
<tr>
<td>T.B</td>
<td>25, m</td>
<td>+100</td>
<td>4/no</td>
<td>3291 ± 1464</td>
<td>3740 ± 1535</td>
</tr>
<tr>
<td>A.B.</td>
<td>29, f</td>
<td>+100</td>
<td>4/no</td>
<td>5026 ± 915</td>
<td>4870 ± 966</td>
</tr>
<tr>
<td>H.K.</td>
<td>26, m</td>
<td>+83</td>
<td>4/no</td>
<td>2226 ± 601</td>
<td>2107 ± 661</td>
</tr>
<tr>
<td>A.P.</td>
<td>27, m</td>
<td>+67</td>
<td>4/no</td>
<td>2090 ± 623</td>
<td>2335 ± 656</td>
</tr>
<tr>
<td>P.R.</td>
<td>24, m</td>
<td>−83</td>
<td>4/no</td>
<td>2116 ± 1094</td>
<td>3046 ± 1248</td>
</tr>
<tr>
<td>M.M.</td>
<td>26, m</td>
<td>+25</td>
<td>4/no</td>
<td>4253 ± 926</td>
<td>4128 ± 938</td>
</tr>
</tbody>
</table>

The handedness score is scaled from −100 (strict left-handedness) to +100 (strict right-handedness). Mean and S.D. values of RTs in the semantic and the color task.
Fig. 1. Sagittal view of slice placement. Midsagittal view of one subject (P.R., also shown in Fig. 5) from Expt. 2. For the actual determination of slice placement during the experiment, additional lateral sagittal images were used.

normally included in the imaged volume. For each plane under investigation $T_1$-weighted spin-echo images ($T_R = 600 \text{ ms}, T_E = 15 \text{ ms}, \text{MA} = 256 \times 256, \text{NEX} = 4$) were acquired from planes identical with those used for functional imaging. Thereafter, strongly $T_2^*$-weighted images were generated using a gradient-echo first-order flow-rephased FLASH pulse sequence ($T_R = 180 \text{ ms}$ (three planes) or $240 \text{ ms}$ (four planes), $T_E = 40 \text{ ms}$, flip angle $a = 40^\circ$, $\text{MA} = 128 \times 128, \text{NEX} = 4$).

We started our experimental procedures by confining ourselves to three slices (Expt. 1), which were oriented such that language-related areas were covered, most notably Broca’s and Wernicke’s areas. A series of FLASH images was measured from three (Expt. 1) or four (Expt. 2) adjacent slices with a temporal resolution of $90 \text{ s}$ (see Fig. 1). In Expt. 1, two activation conditions were interleaved with a resting condition. Each condition was run three times, i.e. the images were obtained under conditions $R/S/C/R/S/C/R/S/C$ ($R$, rest; $S$, semantic judgement; $C$, color judgement), lasting for $\approx 4.5 \text{ min}$ each to produce three images per slice per condition. Expt. 1 lasted for $\approx 70 \text{ min}$, including the anatomical images and shimming procedures. It was carried out in 7 subjects.

10 additional subjects were run under a slightly modified protocol (Expt. 2; reasons for modification will be discussed below), which included the imaging of four (instead of three) slices and the omission of the resting condition. Images were obtained under conditions $S/C/S/C/S/C$, each lasting for $\approx 6 \text{ min}$. The duration of the entire Expt. 2 was $\approx 65 \text{ min}$.

After data acquisition, the matrix size was interpolated to $256 \times 256$ picture elements by zero-filling in the phase-encoding and read-out direction, which led to a nominal in-plane resolution of $0.86 \text{ mm} \times 0.86 \text{ mm}$ in the displayed images. In order to maximize the signal-to-noise ratio of the FLASH images, the acquisition scheme was set up as follows [7]: (1) The flip angle of the excitation pulse was chosen similar to the Ernst angle for gray brain mater ($a_E = \arccos[\exp(-T_R/T_i)] = 35.0^\circ$ (three planes) or $40.0^\circ$ (four planes) for $T_i = 900 \text{ ms}$); (2) the data sample interval was increased to $72 \mu\text{s}$ which allowed bandwidth reduction of the receiver lowpass filter to $13.88 \text{ kHz}$ ($108.5 \text{ Hz/pixel}$); (3) the shape of the water total integral of the free induction decay (FID) signal was optimized by careful magnet shimming in each subject ($\Delta B_0/B_0 \leq 0.2 \text{ ppm}$).

2.4. Data analysis

Behavioral data of each subject were analyzed using customized statistical software. Means and standard deviations of RTs and ERs were computed. ERs were used to check the subjects overall performance of the tasks. Changes of over time in RTs were further analyzed using standard repeated measures analysis of variance (ANOVA) procedures and posthoc comparisons.

Off-line analysis of image data for each subject included the following steps, performed on a DEC 3000/400 AXP workstation (Digital Equipment Corporation, Maynard, MA), in the order presented below, using either in-house developed customized software or standard statistics packages:

1. Residual movement artifacts were reduced by translating and rotating images of a given gradient-echo series such that superposition produced maximal image congruency (two-dimensional affine image transformation).

2. The significance of differences in signal enhancement due to two different stimulus conditions was tested by applying the nonparametric Wilcoxon rank test for different mean values on a pixel-by-pixel basis [44]. The significance level was set to 0.05 in the one-tailed test, as we specifically tested only for positive differences.

3. In order to remove signal contamination due to stochastic noise in isolated pixels, data were further filtered using the following continuity criterion: In a $3 \times 3$ grid around an examined pixel there had to be at least five other significant pixels. It has been shown that such a cluster criterion provides considerable protection against type 1 errors [15,18].

4. The positive relative signal difference in significant pixels was quantified on a scale from 1 to $5\%$. The hereby generated activation maps display the amount of signal increase, which was color coded and overlaid onto the corresponding anatomical $T_1$-weighted spin-echo images.

5. In each slice obtained from each subject, either 8 (frontal, temporal/insula, parietocingual and subcortical) or 10 (frontal, frontalobalateral, M1/S1, parietal and subcortical) anatomically definable regions were delineated manu-
ally using the $T_1$-weighted spin-echo images (Fig. 6). The number of pixels with a significant activation in each of these regions was automatically obtained and divided by the total number of voxels in the region. These normalized activation scores were transferred to a personal computer for further statistical analysis. Data were combined across slices to produce anatomically meaningful activation scores for the following brain regions on the left and right side: frontal, fronto-temporal, parieto-occipital and subcortical.

(6) Hypotheses-driven comparisons between the activation scores in specified regions of interest under the two activation conditions were calculated using Student’s t-test.

(7) The time-course of signal intensity in regions of interest (ROIs) with significant activation due to one of the stimulation conditions was obtained by manually delineating these regions and computerized calculation of the mean activity change within these regions for each image. Time-course data were further processed on a personal computer, and statistical comparisons within an ANOVA framework were used to verify a decrease or an increase in activity over time.

In sum, the dependent variables that entered further statistical analysis consisted of, first, behavioral data (RTs), second, the amount of significant activation in prespecified anatomically defined cortical regions, and third, the change in time-course activation data obtained from manually delineated ROIs.

2.5. Experimental design

Word pairs were created such that the two words were either clearly semantically related (e.g. ‘table-chair’), indirectly semantically related (e.g. ‘beer-grape’) or not related (‘cloud-dog’). These words were typed in a bold face font, one on top of the other, black on a white background. (The word-pairs were taken from [49,50].) For the color judgment task, two strings of four asterisks each were placed in a way similar to the word task on a white background. The colors of the asterisks within the two strings varied. The asterisks within the two strings were either clearly of the same color, or of the same color, but harder to identify as such, or clearly dissimilar in color. In total, 120 word pairs and 120 string-pairs were created such that semantic relation (related: 40; indirectly related: 40; and not related: 40) and color relation (same: 40; same, but harder to detect: 40; and different: 40) were equally distributed regarding task difficulty. Examples of the stimuli are depicted in Fig. 2.

Care was taken to make the two tasks psychometrically equivalent: In a number of preliminary experiments carried out in the psychological laboratory with normal volunteers, the difficulty of the color-task was adjusted such that its difficulty, operationalized in terms of the speed and accuracy of the response, matched the difficulty of the semantic task as closely as possible. One of the reasons to add the intermediate level in both tasks was to make sure that the task was sufficiently difficult for all subjects: When the words are indirectly related, the subjects had to search the associated semantic fields of both of the stimulus words to come up with a common associate. This implies that such a search has to be carried out in the nonassociated condition as well. The net effect of this is that the task strongly involved word association-related mental activity. The intermediate level also obscured the clearly similar cases in adding a certain chance that these cases were in fact instances of dissimilar categories.

The course of the actual experiment was as follows: Subjects were placed in the scanner and first performed two practice experiments with 25 trials each (none of the stimuli of the practice experiments appeared later during
the fMRI experiment). For the fMRI experiment, subjects were instructed to judge word pairs for semantic relatedness, and pairs of asterisks for color similarity; and indicate their decision with their left foot moving to the right for a ‘yes’ and to the left for a ‘no’ response. This response triggered the disappearance of the word pair.

Fig. 4. Typical activation image as obtained in Expt. 1 (subject A.L.). The images in the top row display significantly different pixels in the comparisons between the semantic task and rest (left) and between the color task and rest (right). The positive relative signal difference in significant pixels only was quantified on a color coded scale from 1% (purple) to 5% and more (red). As both tasks are complex, involving perceptual, evaluative, decision-making, and motor components, they activate a network of cortical and subcortical structures. When tasks are subtracted from one another (images on the bottom), components which both tasks have in common are no longer visible. The semantic minus color subtraction (S-C) image (left) shows activation of a left frontal cortical site corresponding to Broca’s area. No significant specific color related activation was detected in the color minus semantic (C-S) subtraction image (right), although in some subjects the C-S comparison showed activation of occipital areas implicated in the visual comparison demanded by the color comparison task.
Fig. 5. Typical activation image as obtained in Expt. 2 (subject P.R.). The two rows correspond to the two activation conditions, semantic (S-C) and color (C-S). The four columns represent the four slices taken. Pixels with significantly ($P < 0.05$) different activity in both activation conditions are depicted and the positive difference in activity (in %) is quantified on a color coded scale from 1% (purple) to 5% and more (red). In the top row, the activation of left frontal and insular regions is most prominent. There is some activity on corresponding regions in the right hemisphere. The depicted example also shows some frontal activity during the color task. This activity, however, was not lateralized to a significant degree in the group comparison, and did not appear consistently across subjects.
followed by a brief blank (the screen went black for one s) to indicate, for the subject and the experimenters, that a response was made. Thereafter, the next stimulus was displayed automatically by the computer.

3. Results

Behavioral data obtained in Expts. 1 and 2 were pooled. They are summarized in Table 1. Mean RTs in the semantic and color condition were not significantly different (2420 ± 985 ms vs. 2527 ± 1002 ms; paired t-test, two-tailed: \( t_{16} = -0.96, P = 0.35 \)). This indicates that the two tasks were psychometrically equivalent in general, i.e. that they were matched with respect to the components of making a decision, and moving the left foot. Presumably, the early stages of their perceptual component can also be regarded as comparable, since the rate of stimulus presentation (which depended upon the subjects pace of doing the task) was not significantly different. Error rates were used only in order to check whether the subjects were able to do the task and in fact performed a decision. They were not subjected to further statistical analysis.

When subjects learn to do a speeded forced-choice task, their RTs usually become shorter. In order to examine whether this was the case, RTs of the three runs of the semantic and of the color task were compared. A two-way-ANOVA with task and number of run as intrasubject factors revealed a significant main effect for run number and a significant interaction (F{\text{2/32}} = 3.87, P = 0.031). When the second and third run were compared to the first run using posthoc comparisons (Sheffe tests), a significant RT decrease was found for the second (\( P < 0.05 \)) and a highly significant decrease was found for the third (\( P < 0.0001 \)) run of the semantic task. Respective comparisons for the color task did not reveal a significant decrease. In other words, run number produced a decrease in RTs, but only in the semantic task (see Fig. 3).

Image data were analyzed by comparing each activation condition with the resting condition (Expt. 1) and by comparing the two activation conditions with each other (Expts. 1 and 2). Inspection of the images obtained in Expt. 1 (cf. Fig. 4) revealed that the comparisons of the two activation conditions (i.e. S-C and C-S) produced more clear cut results than the comparisons of the activation conditions with the resting condition (i.e. S-R and C-R). This was confirmed by two t-tests, the first comparing the amount of activation (i.e. the number of significant pixels in the entire image) in the S-R with the S-C images (\( t_{20} = 4.46, P = 0.0002 \)), and the second comparing activation in the C-R with the C-S images (\( t_{20} = 4.46, P = 0.0002 \)). In both cases, comparing the two tasks produced a highly significant decrease in activation. However, the pattern of activation produced by the task-task comparisons was more interpretable in that activation was limited to a few areas: In the S-C comparison, activation was almost completely restricted to left frontal and left fronto-temporal areas known to be involved in language processing. The C-S comparison showed activation of occipital areas needed for the visual comparison demanded by the color comparison task.

For this reason, the experimental paradigm was changed, with the resting condition left out. The time not spent on obtaining images under the condition of rest was used to obtain data from one more slice (Expt. 2; cf. Fig. 5).

Image data from Expts. 1 and 2 were pooled for further analysis, which first focused on differences in the activation of anatomically specified brain regions by the two conditions S-C and C-S (henceforth referred to as the semantic and the color condition, respectively (cf. Fig. 6). Multiple t-tests between the activation produced by the semantic and the color condition revealed one highly significant difference for the left frontal cortical area (\( t_{60} = 4.36, P < 0.0001 \)) and a significant difference for the left fronto-temporal cortical area (\( t_{60} = 2.06, P < 0.05 \)). In other words, judgements of semantic relatedness between words produced a significant increase in the activation of mainly left frontal areas.

When data from the prespecified regions were averaged across subjects, we did not find activation caused by the color task. However, we detected cortical areas of variable location selectively activated by the color task in all subjects. Finally, it is noteworthy that we found activation of left medial occipital regions caused by the semantic task in some subjects.
In Fig. 7 (top, middle), two examples of the time-course of activation in manually delineated ROIs are shown. These data represent typical activation in two ROIs (C-S and S-C, respectively) from 2 subjects each. Data from all subjects were pooled for each of the two conditions (see Fig. 7, bottom).

The resulting three data points per condition and run were then collapsed. This procedure resulted in three data points per condition per subject, which corresponded to the amount of activity in whatever brain regions were activated in the first, second and third run of the semantic and the color task. Fig. 8 displays the means of these data across all subjects for the semantic and color conditions. Since it was now possible to treat these data similarly to the RT data, a two-way-ANOVA with task and number of run as intrasubject factors was performed, corresponding to the ANOVA on the RTs. A significant main effect for run number and a significant interaction \( F_{3.60, 3.60} = 3.60, P = 0.04 \) were found, and posthoc comparisons (Sheffe tests) revealed highly significant decreases in activation for the second \( P < 0.01 \) and third run \( P = 0.001 \) in the semantic task. Respective comparisons for the color task did not reveal a significant decrease. In other words, a decrease in activation was only detected for the semantic task.

4. Discussion

In this study, we set out to investigate cortical activation during a language-related task. In particular, we wanted to demonstrate the advantages of, first, using psychometrically equivalent control tasks instead of a resting condition, and, second, of relating learning-dependent changes of functional imaging data to behavioral measures obtained during fMRI.

Off-line testing of subjects’ responses in the psychological laboratory verified that the semantic and the color comparison tasks used were of approximately equal difficulty, i.e. were performed by subjects at about the same speed and the same accuracy, assessed as mean RT and ER, respectively, in the semantic and color tasks. Expt. 1
clearly demonstrated the advantage of an active control condition over a rest condition (see Fig. 4). The comparisons of the semantic and the color task, respectively, with the resting condition produce images displaying the activation of a number of cortical sites involved in the performance of both of these rather complex tasks. In contrast, when the activation caused by the color task is subtracted from the activation caused by the semantic task, the main site of activity falls onto the left frontal cortex, and overlaps with Broca’s area.

Although we found specific activation caused by the color task, the color condition did not produce a statistically significant difference across subjects. As the placement of our slices was such that they did not overlap with cortical sites (i.e. the lingual and fusiform gyri of the occipital cortex) previously identified in the processing of color information [10,60,61], it was expected that the color condition did not produce a statistically significant difference across subjects.

Our data strongly suggest that the difference method is applicable to fMRI data, and can be used to identify cortical sites involved in specific mental computations, such as language. Generally speaking, comparisons between similar tasks yield clearer and more readily interpretable results than comparisons between activation conditions and a resting condition (cf. Fig. 4).

This result points to the importance of the design of psychological paradigms for use in functional activation studies. As a rule, active control tasks are better than ill-defined (and never clearly definable) resting conditions. Moreover, control tasks should be designed such that they resemble the ‘target’ task as closely as possible. In the experiments, both conditions involved viewing stimuli of similar size and luminance, performing a cognitive decision-making task of about equal difficulty, and making the same response. Ideally, stimuli could have been constructed identically, with task as the only variable [10,51]. However, numerous psychological studies suggest that words written in colored inks, even when the subjects’ task is to judge color similarity, are nonetheless read and processed semantically [20,26,53]. Though not intended, the psychological processes involved are likely to cause activation of language-related brain areas which would have rendered such a task as not an appropriate control task for the semantic task under scrutiny.

It should be noted that the task used in the experiment encompasses computations similar to the word generation tasks used in several PET studies [35–37] – i.e. it involves the search for meaningfully related word associations of two words – while it does not employ a verbal response. This has two advantages. First, the response time can be measured accurately, which is hardly possible using a voice-activated trigger, given the noise within the operating MR scanner. Second, voiced responses cause activation in motor- and language-related brain regions and introduce head movements, which must by all means be avoided when the fMRI technique is used (see above). We furthermore would like to stress that behavioral data entered our imaging procedures at a rather early stage, i.e. during task construction, when RT and ER data generated in the psychological laboratory guided the selection of words and colors for stimulus presentation.

Pooled data from Expts. 1 and 2 clearly identified the frontal and left fronto-temporal cortical regions as strongly involved in word association-related tasks. This result not only is in line with neuropsychological lesion studies [29], and with intraoperative stimulation studies [32,33], but also with the PET literature on brain activation using word-generation tasks and cognitively similar tasks (for a review, see [42]), and finally, with the few fMRI studies involving language-related tasks [4,5,27,43,48,58]. We furthermore find it noteworthy that a left medial occipital area was found to be activated by the semantic task (invoking reading two words) in some subjects, which is in line with PET data reported by Petersen et al. [37] on a visual word form cortical area.

The ultimate goal of this series of experiments was to relate behavioral data and imaging data. This was done by putting normalized time-course data from ROIs, which were selectively activated either by the semantic or the color task, into a common framework with RTs obtained in these two tasks. The decreased activation of cortical sites caused by practice effects in a language task has recently been demonstrated using PET [41]. Accordingly, we found decreasing RTs over three runs – a behavioral indicator of practice or learning – which were paralleled by decreasing activity in distributed brain regions involved in the performance of the semantic task. It is of critical importance, that this observation was not made in the color task, i.e. that neither behavioral data nor imaging data on a rather similar task, performed three times interleaved with the semantic task, were indicative of the subjects’ increasing ease in performance. In fact, several subjects reported after the experiment that while it was easy to remember some of the words and the judgements on their relatedness, this was not possible for the pairs of colored asterisks. To put it differently, the degree of similarity between colors could not be learned as quickly as the degree of similarity between words. Hence, the small, nonsignificant decrease in RTs in the color task is most likely due to nonspecific learning of the general characteristics of both tasks.

In conclusion, we want to stress two general points pertaining to fMRI research as elucidated by our study. First, it must be emphasized that the fMRI methodology capitalizes upon individual differences rather than group means. This is less obvious (and hence, the point has often been deemphasized) in studies of the primary visual cortex, although even at this area considerable anatomical [54] and certainly no less functional variability can be found. When it comes to language, however, data from electrocortical stimulation studies strongly suggest considerable interindividual variation of functional localization. In partic-
ular, using data from Ojemann et al. [32], Steinmetz and Seitz [52] have convincingly argued that the averaging procedures which are standard in the analysis of PET data result in an underestimation of the number, size and importance of non-Broca cortical areas for language processing. As a corollary, it can be stated that fMRI experiments on language are bound to produce a wider range of activated cortical areas than we are used to see in PET images, and that these apparently more ‘scattered’ results are presumably closer to the truth. This state of affairs calls for alternative procedures of data analysis, which do not rely on between-subject comparisons of normalized coordinates of activated regions in an arbitrary reference brain. In this study, we cautiously approached this problem by deemphasizing localization in two ways: First, we based our statistical comparison of activated brain sites on rather large areas, that were delineated on the grounds of individual anatomy (cf. Fig. 6) rather than by superimposition of a geometrical grid. Second, the link of the time-course data and the behavioral data was obtained by averaging across different regions in different subjects, the ROIs being characterized on purely functional grounds (cf. Fig. 3 and 8). However, this approach does not preclude a more detailed analysis of activation location, which, in our point of view, can be, and should be, accurately performed on an individual basis (cf. Fig. 7).

A second general point directly pertains to the importance of experimental psychology in fMRI. While considerable effort is currently devoted to technical aspects of the imaging procedures, and whereas several investigators have suggested elegant ways of analyzing time-series data of the kind generated in fMRI experiments [1,2], [17,18], comparatively little attention has been paid to the psychometric aspects of fMRI. With methodological aspects of imaging and data analysis more and more worked out, and with the field moving quickly from imaging of primary sensory and motor cortical areas to higher cognitive functions, behavioral and psychometric considerations become increasingly important. Armed with the capability of detecting individual differences in the distributed cortical networks involved in the performance of high-level cognitive tasks, we cannot afford to neglect similar differences in the psychological realm. In brief: The more sophisticated the imaging technology, the more time and effort has to be spent on the assessment of the mental functions that are imaged. The present study may serve as a simple example of how psychometrics and fMRI can be conjointly applied to the study of a variety of higher cognitive functions.

Acknowledgements

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG Sp 364/1-2) and by an established investigator’s award from the National Al-iance for Research in Schizophrenia and Depression (NARSAD) to the first author.

References


