

A Systematic Review and Quantitative Appraisal of fMRI Studies of Verbal Fluency: Role of the Left Inferior Frontal Gyrus

Sergi G. Costafreda,^{1,2*} Cynthia H.Y. Fu,^{2,3} Lucy Lee,⁴ Brian Everitt,¹
Michael J. Brammer,¹ and Anthony S. David²

¹Brain Image Analysis Unit, Department of Biostatistics and Computing, Institute of Psychiatry, King's College London, London, United Kingdom

²Division of Psychological Medicine, Institute of Psychiatry, King's College London, London, United Kingdom

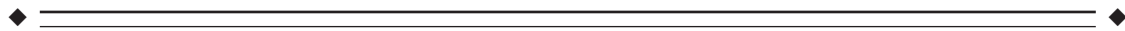
³Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom

⁴Functional Imaging Laboratory, Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, London, United Kingdom



Abstract: The left inferior frontal gyrus (LIFG) has consistently been associated with both phonologic and semantic operations in functional neuroimaging studies. Two main theories have proposed a different functional organization in the LIFG for these processes. One theory suggests an anatomic parcellation of phonologic and semantic operations within the LIFG. An alternative theory proposes that both processes are encompassed within a supramodal executive function in a single region in the LIFG. To test these theories, we carried out a systematic review of functional magnetic resonance imaging studies employing phonologic and semantic verbal fluency tasks. Seventeen articles meeting our pre-established criteria were found, consisting of 22 relevant experiments with 197 healthy subjects and a total of 41 peak activations in the LIFG. We determined 95% confidence intervals of the mean location (x , y , and z coordinates) of peaks of blood oxygenation level-dependent (BOLD) responses from published phonologic and semantic verbal fluency studies using the nonparametric technique of bootstrap analysis. Significant differences were revealed in dorsal–ventral (z -coordinate) localizations of the peak BOLD response: phonologic verbal fluency peak BOLD response was significantly more dorsal to the peak associated with semantic verbal fluency (confidence interval of difference: 1.9–17.4 mm). No significant differences were evident in antero–posterior (x -coordinate) or medial–lateral (y -coordinate) positions. The results support distinct dorsal–ventral locations for phonologic and semantic processes within the LIFG. Current limitations to meta-analytic integration of published functional neuroimaging studies are discussed. *Hum Brain Mapp* 27:799–810, 2006. © 2006 Wiley-Liss, Inc.

Key words: prefrontal cortex; inferior frontal cortex; language; functional magnetic resonance imaging; nonparametric statistics; data pooling; meta-analysis; databases



Contract grant sponsor: Medical Research Council (UK); Contract grant sponsor: Wellcome Trust (UK).

*Correspondence to: Dr. Sergi G. Costafreda, Department of Biostatistics and Computing, Institute of Psychiatry, De Crespigny Park, London SE5 8AF United Kingdom. E-mail: s.costafreda@iop.kcl.ac.uk

Received for publication 7 March 2005; Accepted 24 August 2005

DOI: 10.1002/hbm.20221

Published online 1 March 2006 in Wiley InterScience (www.interscience.wiley.com).

INTRODUCTION

The left inferior frontal gyrus (LIFG) has been linked to language since the advent of neuropsychology [Broca, 1861]. This relationship is supported by neuropsychological [Stuss et al., 1998] and neuroimaging studies spanning nearly two decades [Petersen et al., 1988; for reviews see Bookheimer, 2002; Price, 2000]. Functional neuroimaging studies consistently implicate the LIFG for two essential operations in word comprehension and production: phonology [Demonet et al., 1996; Indefrey and Levelt, 2000], which encompasses processes linked to the sound of words, and semantics [Gabrieli et al., 1998], which involve processes associated with the meaning of words. The functional organization of phonologic and semantic processes within the LIFG, however, remains in dispute.

The functional parcellation theory proposes distinct anatomic regions within the LIFG for the two processes. Phonologic operations are subserved by a more posterior and dorsal region in the LIFG, whereas semantic processes are localized to a more anterior and ventral region [Bookheimer, 2002; Fiez, 1997]. This topographic distinction is supported by a wealth of neuroimaging studies aiming to isolate semantic and phonologic processing [reviewed in: Bookheimer, 2002; Gabrieli et al., 1998; Poldrack et al., 1999]. A functional connectivity study of anterior versus posterior regions in the LIFG during a reading task [Bokde et al., 2001] also showed an anatomic distinction for these processes. When activations associated with semantic and phonologic tasks have been compared in the same subjects, however, these studies consistently have failed to find a clear spatial separation of functions [Chee et al., 1999; Klein et al., 1995; Poldrack et al., 1999; Price et al., 1997; Pugh et al., 1996; Shaywitz et al., 1995; but see McDermott et al., 2003]. To account for the contradictory observations, proponents of the functional parcellation hypothesis suggest that there is some degree of automatic recruitment of phonologic processes in semantic tasks and vice-versa [Bookheimer, 2002; McDermott et al., 2003; Poldrack et al., 1999; Schwartz et al., 2003].

An alternative theory proposes a unitary site in the LIFG that has a more general, supramodal executive function that is a component of both phonologic and semantic processes, but is not phonologic or semantic per se. Thompson-Schill [2003] proposed that the LIFG selects task-relevant information among competing alternatives. This would explain its activation by both semantic and phonologic tasks as these processes usually place high demands on the selection of a particular phonologic or semantic feature, respectively, among many other potentially relevant options [Barde and Thompson-Schill, 2002; Thompson-Schill et al., 1998]. In support, Gold and Buckner [2002] found that the retrieval of both semantic and phonologic information involved the same region that comprised both anterior and posterior areas in the LIFG.

In the present study, we sought to examine whether the published evidence supports either of these accounts. We conducted a systematic review and quantitative analysis of

functional neuroimaging studies of verbal fluency, a classic paradigm of language production in which subjects are asked to generate as many words as possible in a limited time and following specific rules. To avoid heterogeneity associated with different neuroimaging techniques, we only included functional magnetic resonance imaging (fMRI) studies. Phonologic or letter fluency requires the generation of words beginning with a particular letter, such as “f,” whereas semantic or category fluency involves the production of examples of a semantic category, such as “animals” [Lezak, 1995]. These two tasks, although sharing some common cognitive processes, differ significantly in the demands they place on phonologic and semantic processing, respectively.

To compute the confidence intervals associated with each task across studies, we used the quantitative method of bootstrap analysis, a nonparametric technique that is not dependent on any a priori assumptions that may limit other parametric methods of analysis [Efron and Tibshirani, 1993; Manly, 1997]. Previous studies aiming to integrate published functional neuroimaging results have consisted largely of a subjective, visual analysis of the peaks of activations drawn from individual studies, whereas quantitative methods usually have relied on parametric techniques with the assumption that the location of brain processes follow a normal (Gaussian) distribution [Chein et al. 2002; Fox et al., 1997, 1999; Paus, 1996; Turkeltaub et al., 2002]. Although empirical estimates of inter-subject spatial variability support a normal distribution of this parameter [Fox et al., 2001], it is unlikely to be the only source of variation between studies. In the present analysis, we used the bootstrap technique as it entails the least number of assumptions.

MATERIALS AND METHODS

Literature Search

A systematic literature search was conducted to identify the relevant studies. Searches were carried out in the databases MEDLINE, EMBASE, and PSYCINFO from January 1990 to September 2003 using the following search terms: (“Magnetic Resonance” or “MR” or “MRI” or “fMRI” or “neuroimaging”) and (“verbal fluency” or “word generation” or “semantic fluency” or “category fluency” or “phonologic fluency” or “letter fluency”).

This computer search was supplemented by hand searches in the following journals for the same period: *Brain*, *Cerebral Cortex*, *Human Brain Mapping*, *Journal of Cognitive Neuroscience*, *Journal of Neurophysiology*, *Nature Neuroscience*, *NeuroImage*, *Neuron*, *Neuropsychologia*, and *Neuroreport*. The reference lists of studies meeting the selection criteria and pertinent review articles were also searched manually for relevant articles. As many studies presented data on several experiments, we followed the criteria outlined by Fox et al. [1998], who conceptualized quantitative neuroimaging reviews as a three-step process: (1) sampling among articles; (2) among experiments within an article; and (3) among activations within an experiment.

Inclusion Criteria

The inclusion criteria were as follows:

1. Studies published in English in peer-reviewed journals, between January 1990 and September 2003;
2. Use of blood oxygenation level-dependent (BOLD) fMRI neuroimaging technique;
3. The sample consisted of healthy adult subjects;
4. The fMRI data were reported using spatial coordinates and labeled as located in the LIFG, or localized to structures within LIFG, such as Brodmann area 44, 45, or 47;
5. The experimental task was either phonologic (letter) fluency, where the subject is presented with a letter and is requested to generate words beginning with the given letter (e.g., "a") in a limited period of time, or semantic (category) fluency, where the subject produces as many examples as possible from a given category (e.g., "kitchen utensils") in a limited period of time; and
6. The control task was described clearly.

Systematic Review

Sixty-seven articles were reviewed manually. Of these, 13 did not report their results using a system of spatial coordinates (i.e., only through images and anatomical labels), 31 studies used a task that did not meet our criteria (e.g., verb generation, translation, etc.), 4 were positron emission tomography (PET) studies, and 2 studies presented results on a single subject. Seventeen studies met our criteria for inclusion: (1) Abrahams et al., 2003; (2) Brammer et al., 1997; (3) Curtis et al., 1998; (4) Fu et al., 2002; (5) Hutchinson et al., 1999; (6) Knecht et al., 2003; (7) Lurito et al., 2000; (8) Perani et al., 2003; (9) Phelps et al., 1997; (10) Schlösser et al., 1999; (11) Smith et al., 1996; (12) Paulesu et al., 1997; (13) Crosson et al., 1999; (14) Crosson et al., 2001; (15) Gaillard et al., 2003; (16) Gurd et al., 2002; and (17) Pihlajamäki et al., 2000.

The articles described 22 relevant experiments, consisting of 197 healthy subjects and a total of 41 peak activations in the LIFG. The fMRI acquisition parameters, preprocessing and analysis methods, and subject demographics are summarized in the Appendix (Tables 1A and 2A).

All experiments reported at least one activation in LIFG. Experimental details and coordinates of peak activation are presented in Table I. Some studies introduced variations to the standard tasks and therefore each experiment was checked for compatibility with the original definition. This led to the exclusion of one of the three semantic verbal fluency tasks from Crosson et al. [2001]. In this task, subjects were given a category and asked to generate a response at the prompt of a further descriptor; for example, subjects were given the category "birds," followed by the descriptors "flightless...bald..." which might generate the category members "emu...eagle." As well, we excluded a peak activation ($x = -6$, $y = 18$, $z = 50$) from Schlösser et al. [1998], which was deemed to be external to the LIFG [Talairach and Tournoux, 1988].

Other studies modified the tasks in more subtle ways and were not excluded from the analysis. The paradigm of Crosson et al. [1999] entailed the generation of emotional words by giving categories such as desserts (positive connotations) and diseases (negative connotation), whereas emotionally neutral categories were birds or types of rooms. Fu et al. [2002] reported separately trials using "hard" letters (A, G, F, N, E, O, I) and "easy" letters (B, R, L, S, T, P, C), and Schlösser et al. [1999] reported results separately for male and female subjects.

There were 31 activations from 14 experiments that used a phonologic fluency task (mean number of activations per experiment: 2.2), and 10 activations from eight experiments that used a semantic fluency task (mean number of activations per experiment: 1.2). The data have been plotted in Figure 1. A visual inspection does not reveal a clear spatial segregation of the two distributions, although a tendency for semantic activations to be located in more anterior and basal areas of LIFG may be inferred.

Statistical Analysis

To compare the distribution of the retrieved semantic and phonologic activations within the LIFG we constructed $100(1 - \alpha)\%$ confidence intervals for the weighted mean location coordinates along each of the three axes (x , y , z) for both groups, with $\alpha = 0.05$. The weight applied to each individual peak activation was equal to the number of subjects in the experiment. The resulting confidence intervals can be visualized as defining a confidence volume for each set of experiments: semantic and phonologic. We also computed the 95% confidence interval of the difference of mean locations between phonologic and semantic experiments. If the latter confidence interval did not include zero (i.e., no difference), this was accepted as evidence for functional segregation.

We used the bootstrap [Efron and Tibshirani, 1993; Manly, 1997] to construct the confidence intervals. The basic premise to this approach is that in the absence of any information about a population, apart from a random sample of n observations obtained from the population, resampling the obtained sample offers the best possible approximation to what might be obtained if we were to sample the original population again. A bootstrap distribution for a given statistic can then be created by resampling this original sample B number of times and re-computing the statistic in each of these new samples. This bootstrap distribution is used to make inferences about the true value of the parameter in the population. As the resampling process is carried out with replacement, each value in the original sample has an equal probability to appear in the new sample at any given draw, and all the new samples have the original size n . The desired $100(1 - \alpha)\%$ confidence interval for the true value of the statistic is delimited by the $\alpha/2$ and $(1 - \alpha)/2$ percentile values that contain the central $100(1 - \alpha)\%$ of this resampled distribution. With $\alpha = 0.05$, the confidence interval is simply defined by the 2.5 and 97.5 empirical percentiles. In our analyses we have used $B = 1,000$ bootstrap replications. This

TABLE I. Description of the Experiments and Foci of Activations

Active task	Control task	Modality of presentation	Cued/free response (pace in sec if cued)	Overt/covert response	Run duration (s)	Foci of activations in LIFG (x, y, z)
Phonological fluency Abrahams et al., 2003	Word repetition ("rest")	Auditory	Cued (6)	Overt	60	-46, 0, 26 -46, 7, 31 -42, 26, 15 -42, 17, -2 -38, 20, -7 -48, 10, 9 -43, 8, 26
Brammer et al., 1997	Word repetition ("rest")	Auditory	Cued (15)	Covert	30	-40, 8, 26 -46, 8, 37
Curtis et al., 1998	Word repetition ("rest")	Auditory	Cued (3)	Covert	30	
Fu et al., 2002						
Easy	Word repetition ("rest")	Visual	Cued (4)	Overt	28	-51, 3, 20
Hard	Word repetition ("rest")	Visual	Cued (4)	Overt	28	-49, 22, -4 -48, 20, 8
Hutchinson et al., 1999	Forward counting	Auditory	Free	Covert	50	-46, 20, 22
Knecht et al., 2003	Repetition of a memorized letter string ("bababa")	Visual	Free	Covert	15	-50, 6, 26
Lurito et al., 2000	Visual fixation on a letter-like symbol	Visual	Free	Covert	30	-53, 14, 6
Perani et al., 2003 Bilingual	Silent rest	Auditory	Free	Covert	30	-44, 10, 32 -54, 12, 24 -36, 36, 4 -30, 24, 4 -42, 28, 0
Phelps et al., 1997	Word repetition	Auditory	Cued (3)	Overt	48	-46, 24, 18
Schlösser et al., 1999 Males	Forward counting	Auditory	Free	Covert	50	-52, 30, 12 -48, 8, 26 -48, 14, 20
Females	Forward counting	Auditory	Free	Covert	50	-48, 14, 26 -46, 28, 20
Smith et al., 1996	Silent rest	Auditory	Free	Covert	30	-47.5, 30.5, 2.5 -41.5, 38.5, 2.5 -45.5, 11.5, 24.5
Paulesu et al., 1997	Silent rest	Auditory	Free	Covert	30	-36, 24, 16 -36, 6, 20
Semantic fluency Paulesu et al., 1997	Silent rest	Auditory	Free	Covert	30	-36, 24, 16
Crosson et al., 1999 Neutral	Word repetition	Auditory	Free	Covert	18.4	-30, 20, 12 -51, 20, 3
Emotional	Word repetition	Auditory	Free	Covert	18.4	-51, 22, 3
Crosson et al., 2001 Free	Word repetition	Auditory	Free	Covert	17.4	-37, 25, 4
Paced	Word repetition	Auditory	Cued (2.9)	Covert	17.4	-40, 30, 2
Gaillard et al., 2003	Silent rest	Auditory	Free	Covert	32	-48, 22, -6
Gurd et al., 2002	Overlearned sequence (e.g., days of the week)	Auditory	Cued (2)	Covert	30	-36, 4, 28 -36, 22, -10
Pihlajamäki et al., 2000	Forward counting	Visual	Free	Covert		-38, 24, 6

Cued or free response: whether the subjects were prompted to produce a word by an external cue or were free to produce them at their own rhythm. Run duration (s): time in seconds during which the subjects had to produce words for each run of the active task (there were several runs per experiment). LIFG, left inferior frontal gyrus.

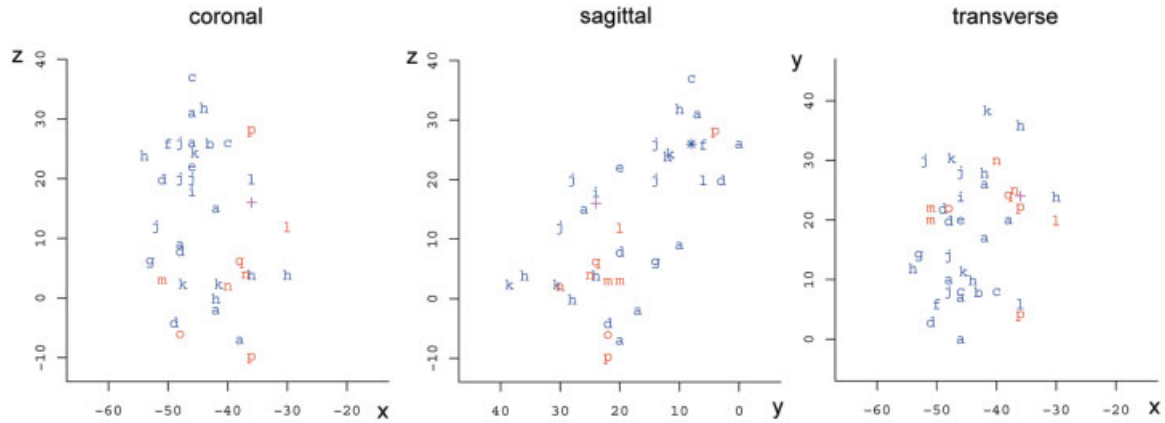


Figure 1.

Peak activations for the studies included in the review plotted on the coronal, sagittal and transverse planes. Individual peaks are represented by blue (phonological) or red letters (semantic). Two overlapping activations from Study I [Paulesu et al., 1997], one corresponding to a phonological experiment and the other to a semantic one, have been replaced by a magenta +. In the sagittal plane, three phonological overlapping peak activations belonging to the Studies b, c, and j have been replaced by a star. The letters

correspond to the studies as reported in Table I. (a) Abrahams et al., 2003; (b) Brammer et al., 1997; (c) Curtis et al., 1998; (d) Fu et al., 2002; (e) Hutchinson et al., 1999; (f) Knecht et al., 2003; (g) Lurito et al., 2000; (h) Perani et al., 2003; (i) Phelps et al., 1997; (j) Schlösser et al., 1998; (k) Smith et al., 1996; (l) Paulesu et al., 1997; (m) Crosson et al., 1999; (n) Crosson et al., 2001; (o) Gaillard et al., 2003; (p) Gurd et al., 2002; and (q) Pihlajamäki et al., 2000.

number of replications usually is accepted as adequate to obtain stable bootstrap confidence intervals [Efron and Tibshirani, 1993].

Two bootstrap analyses were carried out: (1) to determine the spatial (x , y , z) locations of phonological and

semantic verbal fluency tasks in the LIFG, and (2) to assess the impact of the normalisation template on locations of activation with a comparison of the Talairach and Tournoux [1988] and the Montreal Neurological Institute [MNI; Evans et al., 1993] coordinate systems. All compu-

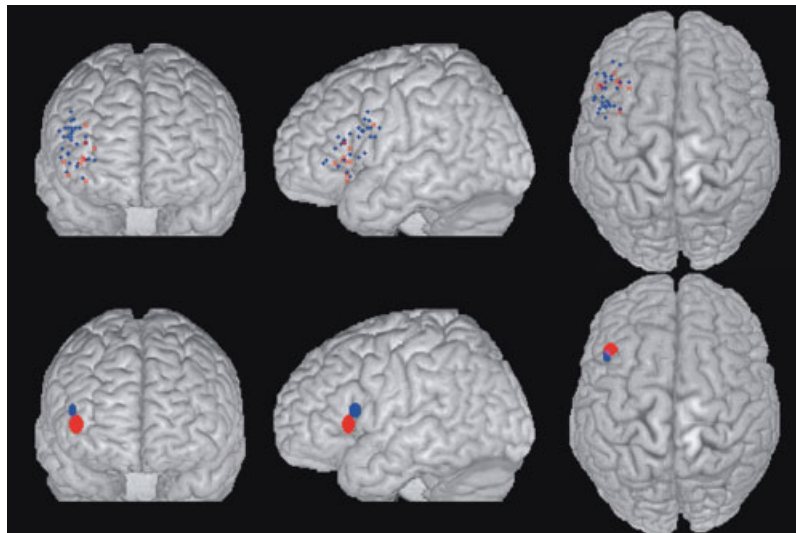


Figure 2.

Peak activations and confidence intervals for phonological and semantic verbal fluency. Coronal, sagittal, and transverse views of a rendered image of the brain (MNI template). Individual peaks are represented as blue crosses (phonological) or red “X” (semantic). Confidence Intervals are represented as blue (phonological) or red (semantic) ellipses. Areas of intersection (phonological \cap semantic) are shown in mauve.

TABLE II. Confidence Intervals of the Phonologic and Semantic Peak Activations and Their Difference for Each Axis

	<i>x</i>	<i>y</i>	<i>z</i>
Phonologic	-46.8, -42.8	13.4, 20.5	10.2, 19.2
Semantic	-46.9, -38.3	17.3, 24.9	-0.5, 11.6
Difference	-7.2, 2.7	-9.3, 1.3	1.9, 17.4

Values shown are 95% confidence intervals. Difference shows the confidence interval of the difference between the location of phonologic and semantic activations, for each axis. It indicates a statistically significant difference at $\alpha = 0.05$ if the confidence interval doesn't include the zero value.

tations were carried out using the statistical analysis language S-Plus.

RESULTS

The confidence interval for BOLD responses associated with semantic verbal fluency showed a position that was more anterior, medial, and ventral to the confidence interval for phonologic verbal fluency (Table II). Some degree of overlap was evident in the three axes. The regional difference between tasks is statistically significant if the confidence interval for the difference does not include zero. No significant differences were found in the antero-posterior (*x*) or medial-lateral (*y*) axes. The confidence intervals for BOLD responses associated with semantic verbal fluency, however, were significantly more ventral (*z*-axis) than were those for phonologic verbal fluency. An overlay of the original peak activations and the confidence intervals on a rendered MNI brain is presented in Figure 2. The additional analysis did not reveal any significant differences in the locations of peak BOLD response according to the template (Talairach or MNI) used for normalization.

DISCUSSION

Spatial Segregation of Phonologic and Semantic Processing in LIFG

The functional parcellation model proposes that a more posterior and dorsal region in the LIFG is associated with phonologic processes whereas an anterior and ventral area is recruited by semantic operations [Fiez, 1997]. Our results largely support this model. The bootstrap analysis revealed that the 95% confidence limits for the location of semantic activations were significantly more ventral to the region associated with phonologic processes. The confidence volume for semantic activations also tended to be in a more anterior region than for phonologic studies, but this difference did not reach significance. The absence of a significant antero-posterior difference may be due to insufficient power, particularly if the antero-posterior difference is smaller than the ventro-dorsal difference; however, it may also reflect a genuine antero-posterior overlap of semantic and phonologic processes in the LIFG.

Phonologic and semantic operations are broad neuropsychological concepts encompassing several functionally heterogeneous subprocesses [Friedman et al., 1998; Indefrey and Levelt, 2000], implemented by distinct brain networks. Recent claims have been made for finer-grained models of parcellation in the LIFG, in which the posterior and dorsal region is subdivided further to accommodate several phonologic subprocesses. Empirical evidence for these models [reviewed in Burton, 2001; Gelfand and Bookheimer, 2003; McDermott et al., 2003] support a ventro-dorsal gradient that is linked to semantic-phonologic operations, respectively, which is consistent with our findings. Moreover, in the present review we observed that each experiment typically reported several peaks of activation in the LIFG, particularly with phonologic fluency (mean activations per experiment: phonologic 2.2; semantic 1.2), which is consistent with a more complex picture.

Letter and category fluency tasks also involve additional cognitive operations that are neither phonologic nor semantic and that elicit activation in the LIFG. In particular, both tasks place demands on verbal working memory, which engages the LIFG [Cabeza and Nyberg, 2000; Collette et al., 2002]. Furthermore, there may be incidental recruitment of phonologic processes during semantic fluency and vice versa. Letter fluency performance is facilitated by automatic activation of some semantic operations [Schwartz et al., 2003], and semantic tasks require the engagement of low-level phonologic processes [Poldrack et al., 1999]. Consequently, the verbal fluency tasks would not totally isolate phonologic and semantic components, and this "noise" would thus weaken the spatial differences in regional activations. The complexity of the component neuropsychological processes may explain why direct comparisons of the verbal fluency tasks [Gourovitch et al., 2000; Mummery et al., 1996], as well as other phonologic and semantic tasks [Chee et al., 1999; Klein et al., 1995; Poldrack et al., 1999; Price et al., 1997; Pugh et al., 1996; Shaywitz et al., 1995; but see McDermott et al., 2003], in the same subjects have failed to observe clear evidence for functional parcellation. If the topographic gradient is such a "noisy" one, a single conventional neuroimaging study may lack the statistical power to detect a difference. In these circumstances, a systematic, quantitative review of the literature would provide sufficient power as evident in the present work.

On the whole, the results support the parcellation model; however, a unitary domain-general model could also account for our findings. If letter and semantic verbal fluency tasks differ systematically on the demands they place in a domain-general mechanism, this could be reflected in differences in the extent of activation within LIFG. An asymmetrical difference in the extent of the BOLD responses for each task thus would result in a difference in the location of their respective peaks.

Methodological Issues

Many reviews of functional neuroimaging studies have used the term "meta-analysis"; however, a fundamental fea-

ture of a meta-analysis is the weighting of the results of individual studies by the inverse of an estimate of their variance (usually the standard error of the point estimate) whereby more precise studies would achieve higher weights [Fleiss, 1993]. As functional neuroimaging studies rarely report a measure of the variance associated with the activation, and to our knowledge it has never been used for the weighting of individual studies, we could not include this measure in our analysis, which precluded the use of the term meta-analysis to describe our study. Reviews that have applied a weighting have been with respect to the number of subjects in the individual studies, with the assumption that larger samples would yield more precise estimates [Fox et al., 1997], and we have used the same strategy in the present review.

Inter-individual variability, however, is unlikely to be the only source of heterogeneity between studies. There are many other likely although largely unexplored factors contributing to systematic differences in the results of studies. Subtle variations in the experimental or control task, systematic differences in the populations from which the groups were selected, and heterogeneity in acquisition equipment and analysis methods may add to inter-study variability. It is worth noting that the latter study-related sources of variability have not yet received much attention in functional neuroimaging literature and may be an important source of variance.

The nonparametric Kolmogorov-Smirnov statistic extended to three-dimensional settings has also been used to evaluate the difference between two distributions of activation data [Duncan and Owen, 2000; Hill et al., 2004; Murphy et al., 2003]. In the present study, we used the bootstrap method as it provided confidence intervals, which aid the visualization of the data.

Two related core assumptions underpin our methodology. The first is task decomposition, whereby complex behavior is deemed to be decomposable in operations, and the second is subtractive reasoning, which proposes that the information on the neural basis of some components of behavior can be studied by subtracting the activity generated by two tasks chosen to isolate the components under study. These assumptions are the basis not only of our method, but also of most of the current research employing neuroimaging techniques. In the light of the overall broad agreement between the results of neuroimaging and brain research in general, these assumptions seem reasonable. In addition, recent single-unit electrophysiological data in monkeys coupled with fMRI measures show a linear link between BOLD responses and neural responses to stimuli [Logothetis, 2001]. This gives some solidity to the subtractive approach by providing empirical evidence of additivity at the neuronal level. In practice subtractive reasoning is not without problems, however, especially when comparing a single active and control task. The results will then depend on the choice of a perfectly chosen control task, which would exactly match the active task except for the component of interest. Although this ideal situation is unlikely to hold for

many studies, and these concerns have been discussed in the literature [Friston et al., 1996; Meegan et al., 2004; Newman et al., 2001], the approach remains popular and has been used by the studies included in our survey. In this context, including studies employing a range of control tasks is desirable for our analysis, as we are ensuring that our results are not dependent on the choice of a particular control task. This in our view reinforces the argument to link those results to the difference between the active (i.e., phonologic and semantic) tasks. The downside is that the use of different control tasks can lead to significant inter-study heterogeneity, which we discuss in the next section.

Limitations

The studies included in our review differed in aspects of their design, methodology, and the population under study.

The verbal fluency experiments used two types of control conditions. The first involved covert or overt repetition of a given word ("rest") or of a familiar sequence (e.g., forward counting), which was used by most of the experiments (10 of 14 phonologic and 6 of 8 semantic fluency experiments). The performance of such standardized language production requires at least some low-level phonologic processing. When subtracted from the experimental tasks, they would eliminate this phonologic activity in the final images, which is most likely localized to the more posterior and dorsal areas of the LIFG [Burton, 2001; Bookheimer, 2002; McDermott et al., 2003]. Consequently, most phonologic experiments may have underestimated the extent of phonologic activity. The second type of control condition was a passive task, such as silent rest or visual fixation. Their effect on the pooled analysis is more difficult to ascertain. There is some evidence, however, that a functionally connected brain network including the LIFG might be associated with resting states [Mazoyer et al., 2001; Shulman et al., 1997; Wicker et al., 2003]. Binder et al. [1999] observed activity in anterior and ventral parts of the LIFG (Brodmann's area 47) during the resting state, which was interpreted as spontaneous semantic retrieval. In this framework, the effects would be the opposite to those of a standardized language production control task. Semantic activity would then be underestimated in semantic experiments, which used a resting state (two of eight semantic experiments) control condition.

Subjects generated covert responses in 15 of 17 studies, whereas overt responses were employed in only two articles that had a phonologic task. Direct comparison of overt and covert responses in a stem completion task showed greater LIFG activation with overt than covert responses, but the location of the peak of activation did not change [Palmer et al., 2001]. The form of verbal output therefore was unlikely to have compromised the validity of our findings.

Studies also chose different brain templates to "normalize" individual anatomic differences. These templates can be divided broadly in Talairach atlas-based templates [Talairach and Tournoux, 1988] and MNI-based templates (Evans et al., 1993). The two types of template differ most notably in the temporal regions [Brett et al., 2001, 2002].

Some transformations have been described to increase comparability of different templates [Brett et al., 2001]. In the present review, both templates were equally used by the phonologic (4 of 12 studies used the MNI template) and semantic (2 of 6 studies used the MNI template) experiments. We could not use these transformations in the present review, as it was not possible to establish with certainty the template used by some of the individual experiments (see Table II). The effect of the different templates would have been a widening of the confidence intervals, however, which would have reduced the power to observe a significant difference. Furthermore, the second analysis did not observe an independent effect of the type of template on the location of activations.

The method of analysis and choice of parameters is another factor that is likely to affect the analysis sensitivity [Hopfinger et al., 2000; Strother et al., 2004], although research in the area is limited. Both semantic and phonologic studies employed a variety of analysis methods (Table 2A), although an important analysis parameter, the size of the smoothing filter, was similar in the studies of phonologic and semantic tasks (mean filter size for phonologic studies: 8.2 mm; for semantic studies: 7.0 mm). Again, our results are unlikely to arise from these choices, as there is no suggestion in the literature of an independent effect on location of activation.

Systematic differences in the groups recruited by the semantic and phonologic experiments could have also introduced bias; however, subjects had similar mean ages (mean age in phonologic experiments: 26 years; in semantic experiments: 31 years) and were predominantly right-handed (see Table II). Female subjects, although in the minority for both phonologic and semantic experiments, were relatively more frequent in the latter. Moreover, the evidence for sexual dimorphism in language processing remains mixed [Baxter et al., 2003; Frost et al., 1999; Rossell et al., 2002].

The best approach to assess possible biases is a quantitative analysis of their contribution to the heterogeneity of results between experiments. For example, stratifying the analysis, i.e., performing separate analysis of phonologic versus semantic activation within the different categories of a covariate, would reveal the effect of each factor on the confidence volumes. Using the bootstrap method, the analysis typically requires samples with a minimum of 10 observations per stratum.

Meta-Analysis of Functional Neuroimaging Data

The use of more powerful meta-analytical techniques is limited at present by the paucity of data in a standard fMRI publication. Usually the only published quantitative data are the coordinates of peak activation (typically the centroid or the most statistically significant voxel in the activation volume) and its associated *P*-value. Occasionally, the cluster volume of activation is also reported. The morphologic data needed to define completely the cluster volumes, however, is retrievable from current analysis programs, as each voxel is identified as activated or not activated. Crucially, pub-

lished data usually lack a measure of their variance (i.e., standard error), which precludes the use of meta-analytical techniques [Fleiss, 1993].

Initiatives such as the pioneering fMRI Data Center (fMRI-DC) at Dartmouth University, [http://www.fmridc.org; Van Horn et al., 2004], or NeuroGenerator [http://www.neurogenerator.org; Roland et al., 2001] are public database of fMRI data, from the “raw” data before any statistical analysis to the final activation maps. They also store the “meta-data” for each study, which is the description of the experimental conditions (from scanner settings to subjects’ characteristics) necessary for the analysis and interpretation of the findings. The statistical brain maps in these databases would be the ideal material for meta-analysis as they could contain all the necessary information about cluster volume, morphology, effect size, and standard error for each voxel. These databases are set to grow as the case for data sharing is convincingly made [Koslow, 2002] and backed increasingly by funding agencies. The development of methods for the pooled statistical analysis of maps from different studies would be an important step in the application of meta-analytical techniques to neuroimaging data.

ACKNOWLEDGMENTS

This work was supported by a Special Training Fellowship in Neuroinformatics from the Medical Research Council UK (to S.G.C.) and by the Wellcome Trust UK (to C.H.Y.F.).

REFERENCES

- Abrahams S, Goldstein LH, Simmons A, Brammer MJ, Williams SC, Giampietro VP, Andrew CM, Leigh PN (2003): Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses. *Hum Brain Mapp* 20:29–40.
- Barde LHF, Thompson-Schill SL (2002): Models of functional organization of lateral prefrontal cortex in verbal working memory: evidence in favor of the process model. *J Cogn Neurosci* 14: 1054–1063.
- Baxter LC, Saykin AJ, Flashman LA, Johnson SC, Guerin SJ, Babcock DR, Wishart HA (2003): Sex differences in semantic language processing: a functional MRI study. *Brain Lang* 84:264–272.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW (1999): Conceptual processing during the conscious resting state: a functional MRI study. *J Cogn Neurosci* 11:80–93.
- Bokde AL, Tagamets MA, Friedman RB, Horwitz B (2001): Functional interactions of the inferior frontal cortex during the processing of words and word-like stimuli. *Neuron* 30:609–617.
- Bookheimer S (2002): Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci* 25:151–188.
- Brammer MJ, Bullmore ET, Simmons A, Williams SC, Grasby PM, Howard RJ, Woodruff PW, Rabe-Hesketh S (1997): Generic brain activation mapping in functional magnetic resonance imaging: a non-parametric approach. *Magn Reson Imaging* 15:763–770.
- Brett M, Lancaster J, Christoff K (2001): Using the MNI brain with the Talairach atlas. *Neuroimage* 13(Suppl):85.
- Brett M, Johnsrude IS, Owen AM (2002): The problem of functional localization in the human brain. *Nat Rev Neurosci* 3:243–249.

- Broca P (1861): Remarques sur le siege de la faculte du langage articule; suivies d'une observation d'aphemie (perte de la parole): *Bull Mem Soc Anat Paris* 36:330–357.
- Burton MW (2001): The role of inferior frontal cortex in phonological processing. *Cogn Sci* 25:695–709
- Cabeza R, Nyberg L (2000): Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J Cogn Neurosci* 12:1–47.
- Chee MW, O'Craven KM, Bergida R, Rosen BR, Savoy RL (1999): Auditory and visual word processing studied with fMRI. *Hum Brain Mapp* 7:15–28.
- Chein JM, Fissell K, Jacobs S, Fiez JA (2002): Functional heterogeneity within Broca's area during verbal working memory. *Physiol Behav* 77:635–639.
- Collette F, Van der Linden M (2002): Brain imaging of the central executive component of working memory. *Neurosci Biobehav Rev* 26:105–125.
- Crosson B, Radonovich K, Sadek JR, Gokcay D, Bauer RM, Fischler IS, Cato MA, Maron L, Auerbach EJ, Browd SR, Briggs RW (1999): Left-hemisphere processing of emotional connotation during word generation. *Neuroreport* 10:2449–2455.
- Crosson B, Sadek JR, Maron L, Gokcay D, Mohr CM, Auerbach EJ, Freeman AJ, Leonard CM, Briggs RW (2001): Relative shift in activity from medial to lateral frontal cortex during internally versus externally guided word generation. *J Cogn Neurosci* 13:272–283.
- Curtis VA, Bullmore ET, Brammer MJ, Wright IC, Williams SC, Morris RG, Sharma TS, Murray RM, McGuire PK (1998): Attenuated frontal activation during a verbal fluency task in patients with schizophrenia. *Am J Psychiatry* 155:1056–1063.
- Demonet JF, Fiez JA, Paulesu E, Petersen SE, Zatorre R (1996): PET studies of phonological processing: a critical reply to Poeppel. *Brain Lang* 55:352–379.
- Duncan J, Owen AM (2000): Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23:475–483.
- Efron B, Tibshirani RJ (1993): An introduction to the bootstrap (1st ed.). New York: Chapman and Hall/CRC. 456 p.
- Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM (1993): 3D statistical neuroanatomical models from 305 MRI volumes. In: Klaisner L, editor. *IEEE Conference Record, Nuclear Science Symposium and Medical Imaging Conference*. San Francisco. p 1813–1817.
- Fiez JA (1997): Phonology, semantics and the role of the left inferior prefrontal cortex. *Hum Brain Mapp* 5:79–83.
- Fleiss JL (1993): The statistical basis of meta-analysis. *Stat Methods Med Res* 2:121–145.
- Fox PT, Huang AY, Parsons LW, Xiong JH, Rainey L, Lancaster JL (1999): Functional volumes modeling: scaling for group size in averaged images. *Hum Brain Mapp* 8:143–150.
- Fox PT, Huang AY, Parsons LW, Xiong JH, Zamarripa F, Rainey L, Lancaster JL (2001): Location-probability profiles for the mouth region of Human primary motor-sensory cortex: model and validation. *Neuroimage* 13:196–209.
- Fox PT, Lancaster JL, Parsons LW, Xiong JH, Zamarripa F (1997): Functional volumes modeling: theory and preliminary assessment. *Hum Brain Mapp* 5:306–311.
- Fox PT, Parsons LM, Lancaster JL (1998): Beyond the single study: function/location metaanalysis in cognitive neuroimaging. *Curr Opin Neurobiol* 8:178–187
- Friedman L, Kenny JT, Wise AL, Wu D, Stuve TA, Miller DA, Jesberger JA, Lewin JS (1998): Brain activation during silent word generation evaluated with functional MRI. *Brain Lang* 64:231–256.
- Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RS, Dolan RJ (1996). The trouble with cognitive subtraction. *Neuroimage* 4: 97–104.
- Frost JA, Binder JR, Springer JA, Hammeke TA, Bellgowan PS, Rao SM, Cox RW (1999): Language processing is strongly left lateralized in both sexes. Evidence from functional MRI. *Brain* 122: 199–208.
- Fu CHY, Morgan K, Suckling J, Williams SC, Andrew C, Vythelingum GN, McGuire PK (2002): A functional magnetic resonance imaging study of overt letter verbal fluency using a clustered acquisition sequence: greater anterior cingulate activation with increased task demand. *Neuroimage* 17:871–879.
- Gabrieli JDE, Poldrack RA, Desmond JE (1998): The role of left prefrontal cortex in language and memory. *Proc Natl Acad Sci USA* 95:906–913.
- Gaillard WD, Sachs BC, Whitnah JR, Ahmad Z, Balsamo LM, Petrella JR, Braniecki SH, McKinney CM, Hunter K, Xu B, Grandin CB (2003): Developmental aspects of language processing: fMRI of verbal fluency in children and adults. *Hum Brain Mapp* 18:176–185.
- Gelfand JR, Bookheimer SY (2003): Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron* 38: 831–842.
- Gold BT, Buckner RL (2002): Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron* 35:8030–812.
- Gourovitch ML, Kirkby BS, Goldberg TE, Weinberger DR, Gold JM, Esposito G, Van Horn JD, Berman KF (2000): A comparison of rCBF patterns during letter and semantic fluency. *Neuropsychology* 14:353–60.
- Gurd JM, Amunts K, Weiss PH, Zafiris O, Zilles K, Marshall JC, Fink GR (2002): Posterior parietal cortex is implicated in continuous switching between verbal fluency tasks: an fMRI study with clinical implications. *Brain* 125:1024–1038.
- Hill K, Mann L, Laws KR, Stephenson CM, Nimmo-Smith I, McKenna PJ (2004): Hypofrontality in schizophrenia: a meta-analysis of functional imaging studies. *Acta Psychiatr Scand* 110:243–56.
- Hopfinger JB, Buchel C, Holmes AP, Friston KJ (2000): A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. *Neuroimage* 11:326–333.
- Hutchinson M, Schiffer W, Joseffer S, Liu A, Schlosser R, Dikshit S, Goldberg E, Brodie JD (1999): Task-specific deactivation patterns in functional magnetic resonance imaging. *Magn Reson Imaging* 17:1427–1436.
- Indefrey P, Levelt WJ. 2000. The neural correlates of language production. In: Gazzaniga M, editor. *The new cognitive neurosciences*. Cambridge, MA: MIT Press. p 845–865.
- Klein D, Milner B, Zatorre RJ, Meyer E, Evans AC (1995): The neural substrates underlying word generation: a bilingual functional-imaging study. *Proc Natl Acad Sci USA* 92:2899–2903.
- Knecht S, Jansen A, Frank A, van Randenborgh J, Sommer J, Kanowski M, Heinze HJ (2003): How atypical is atypical language dominance? *Neuroimage* 18:917–927.
- Koslow S (2002): Sharing primary data, a threat or asset to discovery? *Nat Rev Neurosci* 3:311–313.
- Lezak MD (1995): *Neuropsychological Assessment* (3rd ed.). New York: Oxford University Press. p 544–546.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001): Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157.
- Lurito JT, Kareken DA, Lowe MJ, Chen SH, Mathews VP (2000): Comparison of rhyming and word generation with FMRI. *Hum Brain Mapp* 10:99–106.

- Manly BF (1997): Randomization, bootstrapping and Monte Carlo methods in biology (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC. 424 p.
- Mazoyer B, Zago L, Mellet E, Bricogne S, Etard O, Houde O, Crivello F, Joliot M, Petit L, Tzourio-Mazoyer N (2001): Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res Bull* 54:287–298.
- McDermott KB, Petersen SE, Watson JM, Ojemann JG (2003): A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 41:293–303.
- Meegan DV, Purc-Stephenson R, Honsberger MJM, Topan M (2004) Task analysis complements neuroimaging: an example from working memory research. *Neuroimage* 21:1026–1036.
- Mummery CJ, Patterson K, Hodges JR, Wise JS (1996): Generating ‘tiger’ as an animal name or a word beginning with T: differences in brain activation. *Proc R Soc Lond B Biol Sci* 263:989–995.
- Murphy FC, Nimmo-Smith I, Lawrence AD (2003): Functional neuroanatomy of emotions: a meta-analysis. *Cogn Affect Behav Neurosci* 3:207–233.
- Newman SD, Twieg DB, Carpenter PA (2001): Baseline conditions and subtractive logic in neuroimaging. *Hum Brain Mapp* 14: 228–235.
- Palmer ED, Rosen HJ, Ojemann JG, Buckner RL, Kelley IW, Petersen SE (2001): An event-related fmri study of overt and covert word stem completion. *Neuroimage* 14:182–193.
- Paulesu E, Goldacre B, Scifo P, Cappa SF, Gilardi MC, Castiglioni I, Perani D, Fazio F (1997): Functional heterogeneity of left inferior frontal cortex as revealed by fMRI. *Neuroreport* 8:2011–2017.
- Paus T (1996): Location and function of the human frontal eye-field: a selective review. *Neuropsychologia* 34:475–483.
- Perani D, Abutalebi J, Paulesu E, Brambati S, Scifo P, Cappa SF, Fazio F (2003): The role of age of acquisition, language usage in early, high-proficient bilinguals: an fMRI study during verbal fluency. *Hum Brain Mapp* 19:170–182.
- Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988): Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 362:342–345.
- Phelps EA, Hyder F, Blamire AM, Shulman RG (1997): FMRI of the prefrontal cortex during overt verbal fluency. *Neuroreport* 8:561–565.
- Pihlajamäki M, Tanila H, Hanninen T, Kononen M, Laakso M, Partanen K, Soininen H, Aronen HJ (2000): Verbal fluency activates the left medial temporal lobe: a functional magnetic resonance imaging study. *Ann Neurol* 47:470–476.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999): Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10:15–35.
- Price CJ (2000): The anatomy of language: contributions from functional neuroimaging. *J Anat* 197:225–359.
- Price CJ, Moore CJ, Humphreys GW, Wise RJ (1997): Segregating semantic from phonological processes during reading. *J Cogn Neurosci* 9:727–733.
- Pugh KR, Shaywitz BA, Shaywitz SE, Constable RT, Skudlarski P, Fulbright RK, Bronen RA, Shankweiler DP, Katz L, Fletcher JM, Gore JC (1996): Cerebral organization of component processes in reading. *Brain* 119:1221–1238.
- Roland P, Svensson G, Lindeberg T, Risch T, Baumann P, Dehmel A, Frederiksson J, Halldorsen H, Forsberg L, Young J, Zilles K (2001): A database generator for human brain imaging. *Trends Neurosci* 24:562–564.
- Rossell SL, Bullmore ET, Williams SC, David AS (2002): Sex differences in functional brain activation during a lexical visual field task. *Brain Lang* 80:97–105.
- Schlösser R, Hutchinson M, Joseffer S, Rusinek H, Saarimaki A, Stevenson J, Dewey SL, Brodie JD (1998): Functional magnetic resonance imaging of human brain activity in a verbal fluency task. *J Neurol Neurosurg Psychiatry* 64:492–498.
- Schwartz S, Baldo J, Graves RE, Brugger P (2003): Pervasive influence of semantics in letter and category fluency: a multidimensional approach. *Brain Lang* 87:400–411.
- Shaywitz BA, Pugh KR, Constable RT, Shaywitz SE, Bronen RA, Fulbright RK, Shankweiler DP, Katz L, Fletcher JM, Skudlarski P, Gore JC (1995): Localization of semantic processing using functional magnetic resonance imaging. *Hum Brain Mapp* 2:149–158.
- Shulman GL, Corbetta M, Fiez JA, Buckner RL, Miezin FM, Raichle ME, Petersen SE (1997): Searching for activations that generalize over tasks. *Hum Brain Mapp* 5:317–322.
- Smith CD, Andersen AH, Chen Q, Blonder LX, Kirsch JE, Avison MJ (1996): Cortical activation in confrontation naming. *Neuroreport* 7:781–785.
- Strother S, La Conte S, Kai Hansen L, Anderson J, Zhang J, Pulapura S, Rottenberg J (2004). Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. *Neuroimage* 23(Suppl): 196–207.
- Stuss DT, Alexander MP, Hamer L, Palumbo C, Dempster R, Binns M, Levine B, Izukawa D (1998): The effects of focal anterior and posterior brain lesions on verbal fluency. *J Int Neuropsychol Soc* 3:265–278.
- Talairach J, Tournoux P (1988): Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical. 122 p.
- Thompson-Schill SL, Swick D, Farah MJ, D’Esposito M, Kan IP, Knight RT (1998): Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proc Natl Acad Sci U S A* 95:15855–15860.
- Thompson-Schill SL (2003): Neuroimaging studies of semantic memory: inferring “how” from “where.” *Neuropsychologia* 41: 280–292.
- Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA (2002): Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage* 16:765–780.
- Van Horn JD, Grafton ST, Rockmore D, Gazzaniga MS (2004): Sharing neuroimaging studies of human cognition. *Nat Neurosci* 7:473–481.
- Wicker B, Ruby P, Royet JP, Fonlupt P (2003): A relation between rest and the self in the brain? *Brain Res Brain Res Rev* 43:224–230.

APPENDIX

TABLE 1A. Summary of Studies Included in the Review: Acquisition Variables

Study	Magnet (T)	TE (ms)	TR (ms)	Flip angle (degrees)	In-plane resolution (mm)	Matrix of voxels (mm)	Field of view (mm)	Slice thickness/skip (mm)
Abrahams et al., 2003	1.5	40	6,000	90	—	—	—	7/0.7
Brammer et al., 1997	1.5	40	3,000	90	3 × 3	128 × 64	384 × 192 ^a	7/0.7
Curtis et al., 1998	1.5	40	3,000	—	3 × 3	—	—	7/0.7
Fu et al., 2002	1.5	40	4,000	70	—	—	—	7/1
Hutchinson et al., 1999	1.5	—	5,000	—	1.562 × 1.562 ^a	128 × 128	200	6/1.5
Knecht et al., 2003	1.5	40	2,000	90	3.125 × 3.125	64 × 64	200	7/1
Lurito et al., 2000	1.5	50	2,000	90	3.75 × 3.75 ^a	64 × 64	240	7/2
Perani et al., 2003	1.5	60	3,000	90	4.375 × 4.375	64 × 64	280	4/0
Phelps et al., 1997	2.1	—	—	—	6 × 3 (n = 8) 5 × 2.5 (n = 3)	64 × 64 (n = 8) 32 × 64 (n = 3)	384 × 192 ^a 160 × 160 ^a	5/7
Schlösser et al., 1998	1.5	—	5,000	—	1.562 × 1.562 ^a	128 × 128	200	6/1.5
Smith et al., 1996	—	—	—	40	2 × 2 ^a	128 × 128	256	6/0
Paulesu et al., 1997	1.5	60	3,000	90	2.19 × 2.19	128 × 128	280 × 210	6/0
Crosson et al., 1999	1.5	40	870	45	1.406 × 1.406 ^a	128 × 128	180	6.4-6.9/0
Crosson et al., 2001	1.5	40	870	45	1.406 × 1.406 ^a	128 × 128	180	6.4-6.9/0
Gaillard et al., 2003	1.5	40	4,000	—	3.437 × 3.437 ^a	64 × 64	220	5/0
Gurd et al., 2002	1.5	66	5,000	—	1.15 × 0.898 ^a	200 × 256	230	4/0.3
Pihlajamäki et al., 2000	1.5	70	2,550	90	4.0 × 4.0	64 × 64	256	5/1

TE, echo time; TR, repetition time.

^aEntries computed from other data in the original publication; all other entries are as supplied in the article.

TABLE 2A. Summary of Studies Included in the Review: Preprocessing and Analysis Variables

Study	Template for normalization	Smoothing filter		Analysis software	Voxel-wise probability (P)
		Type	FWHM (mm)		
Abrahams et al., 2003	Talairach	—	—	BAM	0.005
Brammer et al., 1997	Talairach	Gaussian	7	BAM	0.0008
Curtis et al., 1998	Talairach	Gaussian	7	BAM	0.0002
Fu et al., 2002	Talairach	Gaussian	7	BAM	0.0001
Hutchinson et al., 1999	MNI-27	Gaussian	6	SPM 96	0.001
Knecht et al., 2003	MNI-152	Gaussian	6	SMP 99	0.05 (c)
Lurito et al., 2000	Talairach	Hamming	—	—	0.0001
Perani et al., 2003	MNI-152	Gaussian	10	SPM 99	0.001 (c)
Phelps et al., 1997	—	Gaussian	6	—	0.005
Schlösser et al., 1998	MNI-27	Gaussian	15	SPM 96	0.001
Smith et al., 1996	Talairach	—	—	AFNI	—
Paulesu et al., 1997	Talairach	Gaussian	10	SPM 95	0.001
Crosson et al., 1999	Talairach	Gaussian	3	AFNI	0.001
Crosson et al., 2001	Talairach	Gaussian	3	AFNI	0.001
Gaillard et al., 2003	MNI-152	Gaussian	8	SPM 99	0.0001 (c)
Gurd et al., 2002	MNI-152	Gaussian	10	SPM 99	0.05
Pihlajamäki et al., 2000	Talairach	Gaussian	8	MEDx	0.00001

The template for normalization is reported in accordance to the study analysis software. MNI-27 refers to a standard brain from the Montreal Neurological Institute (MNI) created by coregistration of an individual scanned 27 times; MNI-152 refers to another standard created at MNI by averaging the scanners of 152 individuals; Talairach refers to templates self-described as matching the coordinates in the Talairach and Tournoux atlas. BAM refers to the analysis software based on Bullmore et al., 2001. SPM 95, 96, and 99 are the successive versions of the Statistical Parametrical Mapping software (<http://www.fil.ion.ucl.ac.uk/spm>). AFNI is the Analysis of Functional Neuro-Imaging program. MEDx software was developed by Sensor Systems, Sterling, VA. Voxel-wise probability is uncorrected for multiple comparisons except when otherwise stated by marking it (c). FWHM, full width half-maximum.

TABLE 3A. Summary of 17 Studies Included in the Review: Subject Characteristics

Study	Subjects (n)	Age, yr (mean or range)	Handedness (n)	Male/female (n)
Abrahams et al., 2003	18	56.7	18 R	14/4
Brammer et al., 1997	6	—	—	6/0
Curtis et al., 1998	5	31.6	5 R	5/0
Fu et al., 2002	11	30.4	11 R	11/0
Hutchinson et al., 1999	12	23-25	12 R	6/6
Knecht et al., 2003	7	25.7	6 R 1 L	4/3
Lurito et al., 2000	5	27	5 R	2/3
Perani et al., 2003	11	20-27	11 R	11/0
Phelps et al., 1997	11	—	11 R	7/4
Schlösser et al., 1999	12	23	12 R	6/6
Smith et al., 1996	7	34	7 R	2/5
Paulesu et al., 1997	6	28.3	5 R 1 L	6/0
Crosson et al., 1999	17	23.2	17 R	10/7
Crosson et al., 2001	15	23.0	15 R	8/7
Gaillard et al., 2003	29	29.2	29 R	15/14
Gurd et al., 2002	11	32	11 R	6/5
Pihlajamäki et al., 2000	14	26	14 R	7/7