

Dissociation of Mechanisms Underlying Syllogistic Reasoning

Vinod Goel,*† Christian Buchel,* Chris Frith,* and Raymond J. Dolan*‡

*Wellcome Department of Cognitive Neurology, Institute of Neurology, Queens Square, London WC1N 3BG, United Kingdom;

†Department of Psychology, York University, Toronto, Ontario M3J 1P3, Canada; and

‡Royal Free Hospital School of Medicine, Roland Hill St., London NW3, United Kingdom

Received February 28, 2000

A key question for cognitive theories of reasoning is whether logical reasoning is inherently a sentential linguistic process or a process requiring spatial manipulation and search. We addressed this question in an event-related fMRI study of syllogistic reasoning, using sentences with and without semantic content. Our findings indicate involvement of two dissociable networks in deductive reasoning. During content-based reasoning a left hemisphere temporal system was recruited. By contrast, a formally identical reasoning task, which lacked semantic content, activated a parietal system. The two systems share common components in bilateral basal ganglia nuclei, right cerebellum, bilateral fusiform gyri, and left prefrontal cortex. We conclude that syllogistic reasoning is implemented in two distinct systems whose engagement is primarily a function of the presence or absence of semantic content. Furthermore, when a logical argument results in a belief-logic conflict, the nature of the reasoning process is changed by recruitment of the right prefrontal cortex. © 2000 Academic Press

INTRODUCTION

Reasoning is the cognitive activity of evaluating arguments. All arguments involve the claim that one or more propositions (the premises) provide some grounds for accepting another proposition (the conclusion). Deduction is an important form of reasoning which involves explicating information implicit in the premises. Valid deductive arguments involve the claim that their premises provide absolute grounds for accepting the conclusion.

Two theories of deductive reasoning (mental logic and mental models) dominate the cognitive literature. They differ with respect to the competence knowledge they draw upon, the mechanisms they invoke, the mental representations they postulate, and the neuroanatomical predictions they make. Mental logic theories (Braine, 1978; Henle, 1962; Rips, 1994) appeal to an underlying competence knowledge of proof theory (loosely speaking) and a mechanism of inference. This

means that the subject has an underlying knowledge of the inferential role of the closed-form, or logical terms, of the language (e.g., “all,” “some,” “none,” “and,” etc.) and uses this knowledge to infer the conclusion. The claim here is that deductive reasoning is a rule governed syntactic process where internal representations preserve structural properties of linguistic strings in which the premises are stated. This linguistic hypothesis predicts that the neuroanatomical mechanisms of language (syntactic) processing underwrite human reasoning processes.

In contrast, mental model theories (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991) appeal to an underlying competence knowledge of model theory (loosely speaking) and a mechanism of searching through a state space.¹ This means that the subject has an underlying knowledge of the meaning of the closed-form, or logical terms, of the language (e.g., “all,” “some,” “none,” “and,” etc.) and uses this knowledge to construct and search alternative scenarios.² Furthermore, internal representations of the argument are said to preserve the structural properties of the world (e.g., spatial relations) that the sentences are about (rather than the structural properties of the sentences themselves, as above). The claim is that deductive reasoning is a process requiring spatial manipulation and search. Mental model theory is often referred to as a spatial hypothesis and predicts that the neural structures for visuo-spatial processing contribute the basic representational building-blocks used for logical reasoning (Johnson-Laird, 1994).

There is, however, a third alternative. The dual mechanism theory (Evans *et al.*, 1983; Goel, 1995; Slovic, 1996) makes a distinction between form- and content-based reasoning and predicts the involvement of two different mechanisms in human reasoning. Sub-

¹ See Newell (1980) for a discussion of the relationship between search and inference.

² Whether there is any substantive difference between “knowing the inferential role” and “knowing the meaning” of the closed-form terms, and thus the two theories is a moot point, debated in the literature.

jects may reason by either attending to the formal/logical structure of the argument or they may engage a content-sensitive associative mechanism. This is, however, a minority position.

To test these competing hypotheses we used an experimental design that manipulated the content in syllogistic reasoning tasks. The content condition contained arguments such as “All dogs are pets; All pets are furry \therefore All dogs are furry,” while the no-content condition contained logically equivalent arguments of the form “All A are B; All B are C \therefore All A are C.” We scanned 11 right-handed normal subjects using event-related fMRI, to index task related neural activity, while they judged the validity of such arguments. The mental logic theory predicts that both conditions should activate the same language (syntax) processing network. The mental model theory predicts that both conditions should activate the same visuospatial processing network. The dual mechanism theory predicts the involvement of two distinct networks, a semantic/associative network, and a formal reasoning network (be it syntactic or visuospatial).

Our findings demonstrate the engagement of both linguistic and spatial mechanisms in reasoning processes. An important factor in the determination of which mechanism is engaged is the presence or absence of content words in the argument. Furthermore, when a logical argument results in a belief-logic conflict, the nature of the reasoning process is changed by recruitment of the right prefrontal cortex to complete the task.

METHOD

Subjects

Eleven right-handed normal subjects (seven males and four females), with a mean age of 29.4 years (SD 7.2) and mean education level of 17.4 years (SD 2.8), volunteered to participate in the study. All subjects gave informed consent and the study was approved by the Joint National Hospital for Neurology and Neurosurgery/Institute of Neurology Ethics Committee.

Stimuli

Thirty contentful and thirty no-content syllogisms, encompassing 14 different logical forms were generated. Half of the arguments were valid, the other half invalid. The contentful syllogisms contained sentences like “All swans are black” while the no-content sentences were of the form “All P are B.” Within the content arguments, half were congruent, half incongruent. In the congruent form the truth value of the conclusion was consistent with the validity judgment (i.e., if the argument was valid, then the conclusion was true; if the argument was invalid, then the conclu-

Reasoning Condition	Content		No-content
	Incongruent	Congruent	
Reasoning Condition	All pets are poodles All poodles are vicious All pets are vicious	All poodles are pets All pets have names All poodles have names	All P are B All B are C All P are C
Preparation Condition	All dogs are pets All poodles are dogs No napkins are white		All P are B All B are C No D are F
Low level Baseline	All dogs are pets All barbers are tailors No primates are Chimps		All P are B All X are Y No D are F

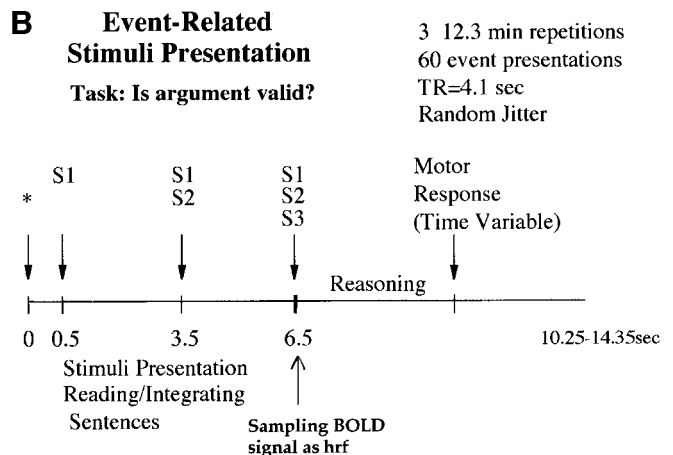


FIG. 1. (A) Experimental design and sample stimuli. (B) Stimuli from all conditions were presented randomly in an event-related design. An * indicated the start of a trial at 0 s. The sentences appeared on the screen one at a time with the first sentence appearing at 500 ms, the second at 3500 ms, and the last sentence at 6500 ms. The length of trials varied from 10.25–14.35 s, leaving subjects 3.75 to 7.85 s to respond. Sampled data was time-locked to the presentation of the third sentence.

sion was false). In the incongruent arguments, the truth value of the conclusion conflicted with the validity of the argument (i.e., if the argument was valid, then the conclusion was false; if the argument was invalid, then the conclusion was true). The 14 argument forms were selected from the easy end of the spectrum of difficulty (Dickstein, 1978). All contentful sentences were grammatical, meaningful, and of roughly equal length.

We used two control conditions, a “preparation” condition, and a low-level baseline. These were created in the following way: The conclusions of the arguments were switched around to generate the preparation condition. The last two sentences were individually switched around to generate the low-level baseline. The latter was used to determine the direction of differential task responses. Examples of each of these category of stimuli appear in Fig. 1A.

Stimuli presentation. Stimuli from all conditions were presented randomly in an event-related design

(see Fig. 1B). The beginning of a trial was signaled by an *. The sentences appeared on the screen one at a time with the first sentence appearing at 500 ms, the second at 3500 ms, and the last sentence at 6500 ms. All sentences remained on the screen until the end of the trial. The length of trials varied from 10.25–14.35 s, leaving subjects 3.75 to 7.85 s (after the presentation of the third sentence) to respond.

Task. The task in *all* trials was the same. Subjects were required to determine whether the given conclusion followed logically from the premises (i.e., whether the argument was valid). In the low-level baseline trials, where all three sentences were unrelated, subjects could disengage after the presentation of the second sentence and respond “no.”³ In preparation trials, where the first two sentences were related, subjects would need to begin to integrate the premises and construct a representation of the problem,⁴ but when the third, unrelated sentence appeared they could immediately disengage the task and respond “no.” In trials where the three sentences constituted an argument, subjects would need to continue with the reasoning component of the task after the presentation of the third sentence (reasoning condition). The difference between completing the reasoning task and disengaging after the presentation of the third sentence will isolate the reasoning components of interest. This design, involving a time-locked single-event design and an (unblocked) random presentation of trials, circumvents the question about what constitutes an appropriate baseline for reasoning tasks and allows us to keep task instructions constant across all conditions.

Subjects responded by pressing a button on a keypad after the appearance of the last sentence. Subjects were instructed to respond as quickly as possible and move to the next trial if the stimuli advanced before they could respond. Subjects reviewed example stimuli from each condition prior to being scanned to ensure that they understood the task.

fMRI Scanning Technique

A 2T Siemens VISION system (Siemens, Erlangen, Germany) was used to acquire T1 anatomical volume images ($1 \times 1 \times 1.5$ -mm voxels) and 48 T2*-weighted echoplanar images (64×64 , 3×3 -mm pixels, TE = 40 ms) sensitive to blood oxygenation level dependent (BOLD) contrast. Echoplanar images (1.8-mm-thick) were acquired axially every 3 mm, positioned to cover

the whole brain. Data were recorded during a single acquisition period. A total of 558 vol images were acquired over three sessions (186 vol per session) with a repetition time (TR) of 4.1 s/vol. The first six volumes in each session were discarded (leaving 180 vol per session) to allow for T1 equilibration effects.

Trials from all conditions were randomly presented in a single-event design. The mean trial time was $12300 \text{ ms} \pm 2050 \text{ ms}$ (TR) with a random jitter. Trials thus varied from 10.25 to 14.35 s. There were 60 event presentations during a session for a total of 180 over the three sessions. Each session lasted 12.3 min. The scanner was synchronized with the presentation of all trials in each session.

Data Analysis

Data were analyzed using Statistical Parametric Mapping (SPM 97d) (Friston *et al.*, 1995). All volumes in a session were spatially realigned to the first volume of the session and temporally realigned to the AC–PC slice, to account for different sampling times of different slices. Subjects with head movement greater than 2 mm were discarded. A mean image created from the realigned volumes was coregistered with the structural T1 volume. The structural volumes were then spatially normalized to the Montreal Neurological Institute brain template (Evans *et al.*, 1993), using nonlinear basis functions (Ashburner and Friston, 1999). The derived spatial transformation was then applied to the realigned T2* volumes, which were finally spatially smoothed with a 10-mm FWHM isotropic Gaussian kernel, in order to make comparisons across subjects and to permit application of random field theory for corrected statistical inference (Worsley and Friston, 1995). The resulting time series across each voxel were high-pass filtered with a cut-off of 120 s, using cosine functions to remove section-specific low frequency drifts in the BOLD signal. Global means were normalized by proportional scaling to a Grand Mean of 100, and the time series temporally smoothed with a 4-s FWHM Gaussian kernel to swamp small temporal autocorrelations with a known filter.

Condition effects at each voxel were estimated according to the general linear model and regionally specific effects compared using linear contrasts. Each contrast produced a statistical parametric map of the t statistic for each voxel, which was subsequently transformed to a unit normal Z distribution. The activations reported survived either a cluster-level or voxel-level correction of $P < 0.05$ ($Z > 4.60$) using a random effect model. An exception was made in the case of interactions. An uncorrected threshold of $P < 0.001$ was used for interactions as they were inclusively masked by the main effect (approximating the joint probability of both effects). In the analysis, the presentations of the first two sentences were modeled out. The sampled data

³ This does mean that there is an imbalance between “yes” and “no” responses. However, the behavioral data suggests that subjects are not getting locked into a mental set. We have explored a number of baselines and feel the advantages of this one outweigh the disadvantages.

⁴ Task difficulty and time limitations do not allow subjects the option of waiting until the presentation of the third sentence before deciding to begin integration of the first two sentences.

TABLE 1

A	Behavioral Scores			
	RTs (ms)		Scores (%)	
	Content	No-content	Content	No-content
Reasoning	3223 (1386)	2986 (1093)	79.0 (32.8)	77.3 (28.7)
Preparation	1641 (529)	1468 (567)	95.4 (9.2)	95.9 (9.0)
Baseline	1447 (583)	1307 (509)	96.7 (8.9)	97.4 (7.3)

B	Congruent		Incongruent	
	RTs (ms)		Scores (%)	
	3115 (1120)		74.1 (35.9)	
RTs (ms)	3115 (1120)		3332 (1564)	
Scores (%)	84.0 (27.6)		74.1 (35.9)	

Note. SD shown in parentheses.

was time-locked to the appearance of the third sentence.

RESULTS

Behavioral scores indicated that subjects performed the task in the expected manner (see Table 1). Subjects took a mean of 3223 ms (SD = 1386) (after presentation of third sentence at 6500 ms) to respond to the content arguments and 2986 ms (SD = 1093) to respond to the no-content arguments. The difference did not reach statistical significance. Performance scores in the two conditions were quite similar (79% (SD =

0.33) for content and 77% (SD = 0.29) for no-content). The mean reaction times for the preparation condition were significantly lower (1555 ms) (SD = 548), and performance scores significantly higher (95.5%) (SD = 0.09), with no difference between content and no-content conditions. The mean reaction time for the low-level baseline condition was 1377 ms (SD = 546), while the performance score was 97% (SD = 0.08), again with no significant difference between content and no-content conditions. Within the content condition, the incongruent trials took a mean of 3332 ms (SD = 1564) vs 3115 ms (SD = 1120) for the congruent trials. This difference approaches but does not reach significance. Performance scores on the congruent trials (84%) (SD = 27.2) were significantly higher ($t = (2.79)$, $P < 0.05$) than scores on incongruent (74.1%) (SD = 35.9).

The main effect of reasoning, derived from comparisons of content and no-content reasoning tasks [(content reasoning + no-content reasoning) – (content preparation + no-content preparation)], with the preparation condition as reference, revealed activation of bilateral cerebellum ($R > L$), bilateral fusiform gyrus, left superior parietal lobe, left middle temporal gyrus, bilateral inferior frontal gyrus, bilateral basal gangli (encompassing the accumbens, caudate nucleus, and putamen), and brain stem (Fig. 2A and Table 2).

However, examination of simple main effects of reasoning revealed two distinct networks. The content condition (content reasoning – content preparation)

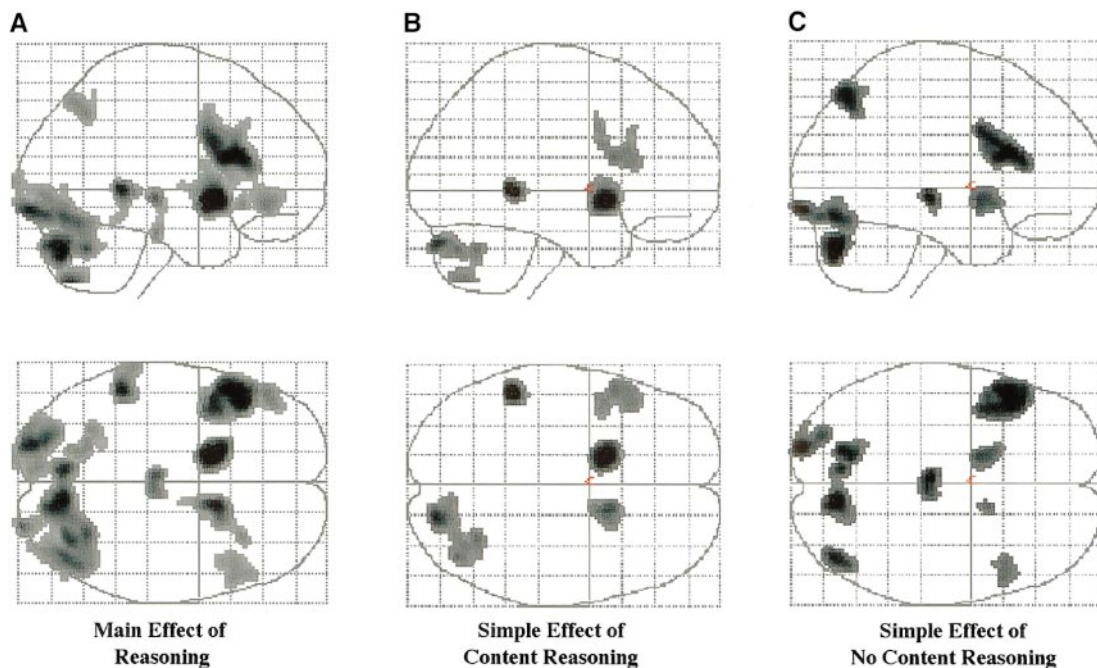


FIG. 2. Sagittal and transverse views of a statistical parametric map (SPM) rendered into standard stereotactic space and projected onto a glass brain. (A) Main effect of reasoning [(content reasoning + no-content reasoning) – (content preparation + no-content preparation)]. (B) Simple effect of content reasoning (content reasoning – content preparation). (C) Simple effect of no-content reasoning (no-content reasoning – no-content preparation). See text.

TABLE 2

Location and Characteristics of the Brain Regions That Remained Significantly Active after Each Subtraction

Location (Brodmann area)	MNI coordinates			Z score
	X	Y	Z	
Main effect of reasoning				
Lt. basal ganglia (accumbens, caudate nucleus, and putamen)	-16	8	-4	7.00
Rt. basal ganglia (accumbens, caudate nucleus, and putamen)	14	10	-4	6.48
Brain stem	0	-22	-4	4.94
Rt. cerebellum	10	-76	-34	6.63
Lt. cerebellum	-6	-74	-30	6.02
Rt. fusiform gyrus (18, 19)	42	-74	-16	5.62
Lt. fusiform gyrus (18)	-20	-92	-12	6.16
Lt. inferior frontal lobe (44)	-52	16	18	6.57
Lt. inferior frontal lobe (44)	-42	10	28	6.09
Lt. inferior frontal lobe (45)	-46	24	20	6.67
Lt. middle temporal lobe (21/22)	-52	-42	0	6.35
Lt. superior parietal lobe (7)	-18	-64	46	4.62
Rt. inferior frontal lobe (45)	52	20	22	4.45
Content – preparation				
Rt. basal ganglia (accumbens, caudate nucleus, and putamen)	14	12	-4	5.21
Lt. basal ganglia (accumbens, caudate nucleus, and putamen)	-16	10	-4	6.10
Rt. cerebellum	20	-84	-30	5.38
Rt. cerebellum	32	-62	-32	4.70
Rt. cerebellum	40	-72	-32	4.67
Lt. inferior frontal lobe (44)	-52	14	20	4.71
Lt. middle temporal lobe (21/22)	-50	-42	2	6.20
No-content – preparation				
Rt. cerebellum	10	-74	-34	5.55
Lt. cerebellum	-8	-72	-28	4.92
Rt. fusiform gyrus (18)	42	-76	-14	4.97
Lt. basal ganglia (accumbens, caudate nucleus, and putamen)	-16	4	-6	4.62
Rt. basal ganglia (accumbens, caudate nucleus, and putamen)	14	10	-4	4.31
Brain stem	-4	-22	-4	5.17
Lt. fusiform gyrus (18)	-20	-94	-10	5.20
Lt. inferior frontal lobe (45)	-44	24	20	5.34
Lt. inferior frontal lobe (44)	-42	14	26	4.89
Lt. superior parietal lobe (7)	-18	-70	52	5.09
Rt. inferior frontal lobe (45)	46	20	22	4.23
Conjunction of content and no-content				
Rt. cerebellum	16	-82	-30	6.04
Lt. basal ganglia (accumbens, caudate nucleus, and putamen)	-16	8	-4	7.07
Rt. basal ganglia (accumbens, caudate nucleus, and putamen)	14	10	-4	6.48
Lt. inferior frontal lobe (44)	-52	14	20	6.32
Lt. fusiform gyrus (18)	-20	-88	-12	5.68
Rt. fusiform gyrus (37)	48	-66	-18	5.15
Content – no-content				
Lt. middle/superior temporal lobe (21/22)	-58	-40	-2	7.66
Lt. inferior frontal lobe (47)	-52	28	-4	7.74
Lt. striate cortex (17) & lingual gyri (18)	-8	-78	-12	9.30
Rt. striate cortex (17) & lingual gyri (18)	12	-76	-10	9.07
	-8	-72	6	
No-content – content				
Lt. occipital gyrus (18, 19)	-48	-72	-2	7.00
Rt. occipital gyrus (18, 19)	50	-70	-6	7.90
Lt. superior parietal lobe (7)	-20	-62	42	7.00
Lt. inferior parietal lobe (40)	-42	-34	40	7.90
Rt. superior parietal lobe (7)	26	-66	50	7.30
Rt. inferior parietal lobe (40)	42	-40	44	7.40
Lt. middle frontal gyrus (6)	-28	2	54	6.40
Rt. middle frontal gyrus (6)	26	6	56	5.20
Lt. precentral gyrus (6)	-56	6	18	6.67
Rt. precentral gyrus (6)	56	8	26	5.16
(Content – preparation) – (no-content – preparation)				
Lt. superior temporal sulcus (21/22)	-54	-40	2	3.75
(No-content – preparation) – (content – preparation)				
Lt. superior parietal lobe (7)	-20	-72	52	4.21
Incongruent – congruent				
Rt. superior frontal gyrus (8)	10	30	54	4.36
Rt. lateral middle/inferior frontal gyrus (46/45)	54	28	26	4.30

resulted in activation of the left middle/superior temporal lobe (BA 21/22), and left inferior prefrontal cortex (BA 44, 45) right cerebellum, and bilateral basal ganglia (Fig. 2B and Table 2). The no-content condition (no-content reasoning – preparation) activated bilateral fusiform gyrus (BA 18), left superior parietal lobe (BA 7), left inferior frontal lobe (BA 44, 45), right inferior frontal lobe (BA 45), bilateral cerebellum ($R > L$), bilateral basal ganglia ($L > R$) (centered around the accumbens, caudate nucleus, and putamen), and brain stem (Fig. 2C and Table 2).⁵

To determine areas common to content and no-content reasoning, we performed a conjunction analysis of the simple main effects of reasoning [Conjunction (content reasoning – content preparation), (no-content reasoning – no-content preparation)]. Common areas of activation were left inferior prefrontal cortex (BA 44), left fusiform gyrus (BA 18), right fusiform gyrus (BA 37), bilateral basal ganglia nuclei (accumbens, caudate nucleus, and putamen), and right cerebellum (Table 2).

The question of distinct activations in the reasoning task as a function of content was determined by direct comparisons of content and no-content conditions (excluding preparation and masked by the main effect of reasoning relative to low-level baseline). The content reasoning – no-content reasoning comparison revealed activation of the left middle/superior temporal lobe (BA 21/22), the left inferior frontal lobe (BA 47), and bilateral (BA 17) and lingual gyri (BA 18) (Figs. 3A–3C and Table 2). The reverse comparison (no-content – content) revealed activation of bilateral occipital (BA 18, 19), bilateral superior and inferior parietal lobes (BA 7, 40), bilateral precentral gyrus (BA 6), bilateral middle frontal gyrus (BA 6) (Figs. 4A–4C and Table 2).

⁵ Integrating content premises, over and above reading unrelated sentences (content preparation – content low-level baseline) resulted in activation of a network involving left superior parietal lobe (BA 7), left middle/superior temporal lobe (BA 21/22), and left inferior and dorsal lateral prefrontal cortex (BA 45, 47, 8). Integrating no-content premises, over and above reading unrelated sentences (no-content preparation – no-content low-level baseline) activated a network involving bilateral occipital lobes (BA 19), bilateral superior parietal lobe (BA 7), bilateral frontal middle gyrus (BA 6), left inferior and dorsal lateral prefrontal cortex (BA 44, 8, 10), and right inferior/middle prefrontal cortex (BA 46).

We next examined the interaction between content and reasoning (masked by the main effect of reasoning relative to low-level baseline). Modulation of the reasoning task by removing content [(no-content reasoning – no-content preparation) – (content reasoning – content preparation)] revealed activation in the left superior parietal lobe (BA 7) unique to the no-content condition (Fig. 5A and Table 2). The reverse modulation of reasoning, by the addition of content [(content reasoning – content preparation) – (no-content reasoning – no-content preparation)] revealed left superior temporal sulcus (BA 21/22) activation unique to the content condition (Fig. 5B and Table 2).

Finally, a direct comparison of the congruent condition with the incongruent condition (congruent – incongruent content) resulted in activation of the superior frontal gyrus (BA 8) (predominately RH) and the right lateral inferior/middle frontal gyrus (BA 46/45) (Fig. 6 and Table 2). The reverse comparison did not reveal any regions of significant activation.

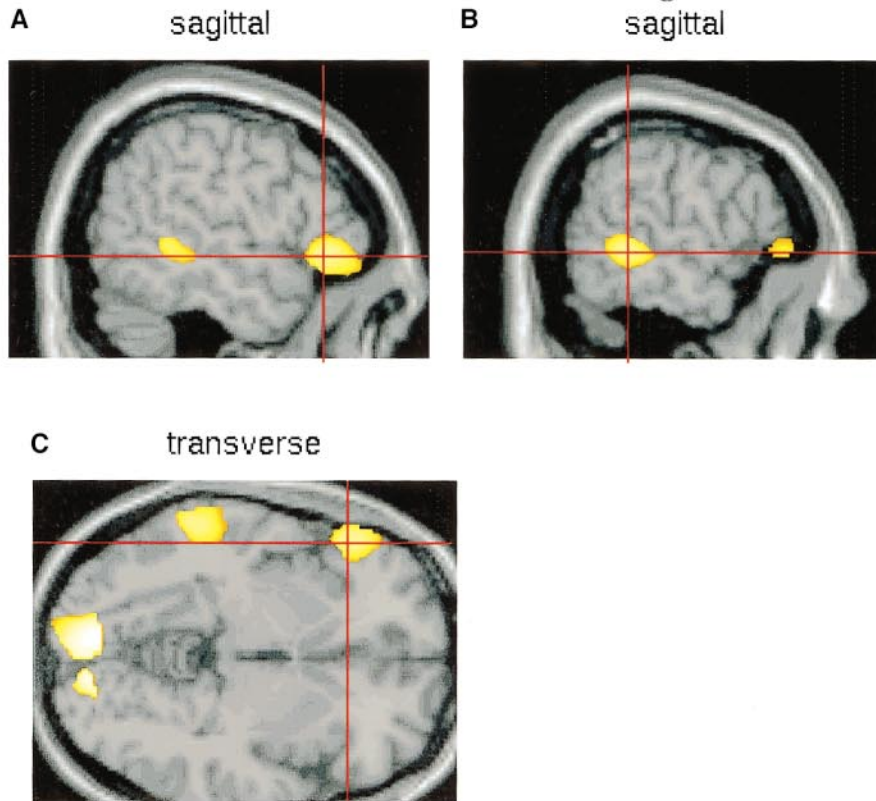
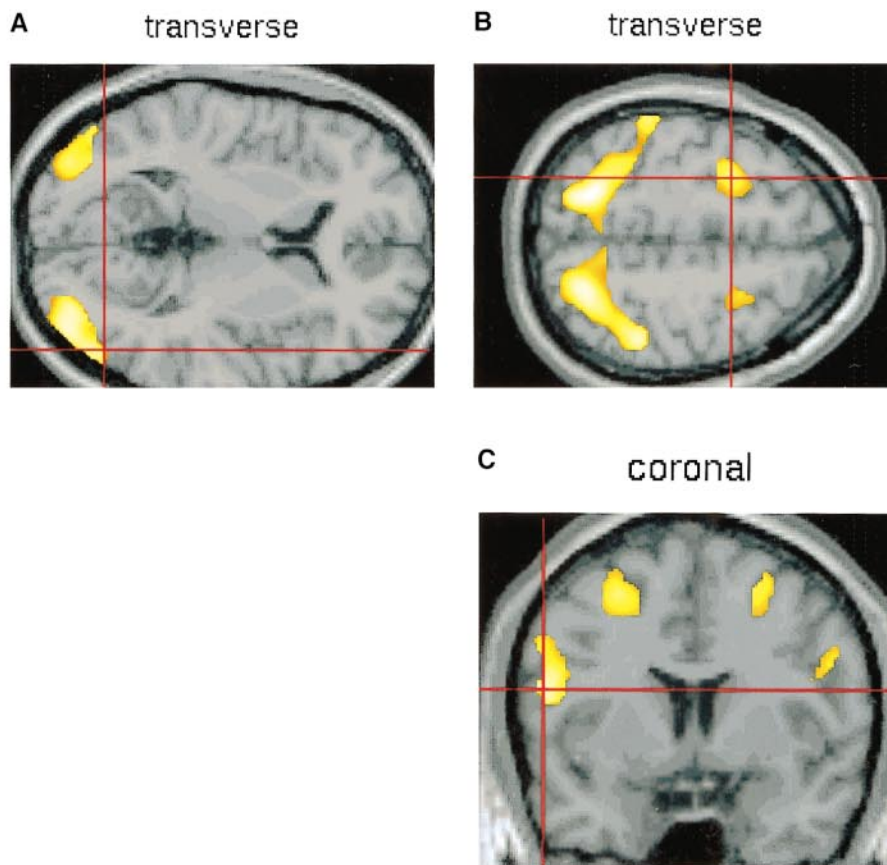
DISCUSSION

Syllogistic reasoning implicates a widespread network involving occipital, temporal, and parietal lobes, prefrontal cortex, and, surprisingly, cerebellum and basal ganglia nuclei. Our network encompasses most the regions found in previous imaging studies of deduction (Goel *et al.*, 1997, 1998; Osherson *et al.*, 1998). However, closer analysis reveals two dissociable anatomical networks for reasoning, with shared common circuits in basal ganglia nuclei, cerebellum, fusiform gyri, and left prefrontal cortical regions.⁶ This dissociation has implications for cognitive theories of reasoning.

⁶ It is becoming increasingly clear from imaging studies (Rao *et al.*, 1997) and patient studies (Channon *et al.*, 1993; Fiez *et al.*, 1992; Grafman *et al.*, 1992; Sandson *et al.*, 1991) that the basal ganglia nuclei and cerebellum have a critical role to play in higher-level cognition. A circuit involving the dorsolateral prefrontal cortex, caudate nucleus, and thalamus have been implicated in working memory and rule-based learning tasks (Cummins, 1993). Anatomical connections have been demonstrated between dorsolateral prefrontal cortex and lateral cerebellum, via the dentate nucleus and mediodorsal nucleus of the thalamus (Middleton and Strick, 1994).

FIG. 3. A statistical parametric map (SPM) rendered into standard stereotactic space and superimposed on to sagittal (A and B) and transverse (C) sections of an magnetic resonance image (MRI), which is itself in standard space. Regions specific to content reasoning (content reasoning – no-content reasoning) include left inferior frontal lobe (–52, 28, –4), left temporal pole (–58, 10, –20), and left middle and superior temporal gyrus (–58, –40, –2).

FIG. 4. A statistical parametric map (SPM) rendered into standard stereotactic space and superimposed on to transverse (A and B) and coronal (C) sections of an magnetic resonance image (MRI), which is itself in standard space. Regions specific to no-content reasoning (no-content reasoning – content reasoning) include (A) bilateral orbital gyrus (–48, –72, –2 & 50, –70, –6); (B) bilateral superior (–20, –62, 42 & 26, –66, 50) and inferior (–42, –34, 40 & 42, –40, 44) parietal and bilateral middle (–28, 2, 54 & 26, 6, 56) frontal gyrus and (C) bilateral inferior (–56, 6, 18 & 56, 8, 26) frontal gyrus is specific to no-content reasoning.

3 Content - No-Content Reasoning**4 No-Content - Content Reasoning**

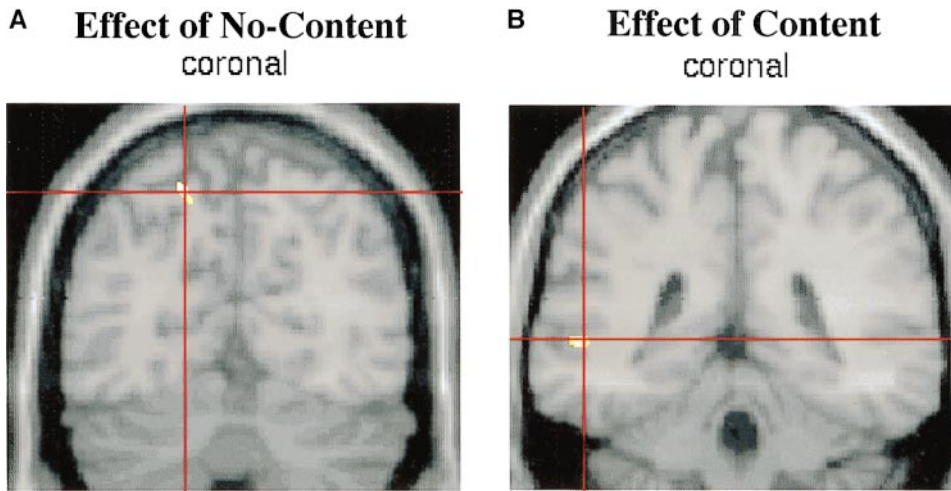


FIG. 5. SPMs rendered into standard stereotactic space and superimposed on to a coronal MRI in standard space. (A) The interaction analysis [(no-content reasoning – preparation) – (content reasoning – preparation)] reveals that the left superior parietal (–20, –72, 52) plays a critical role in reasoning in the absence of content. (B) The reverse interaction [(content reasoning – preparation) – (no-content reasoning – preparation)] reveals that the left superior temporal sulcus (–54, –40, 2) plays a critical role in reasoning in the presence of content.

Reasoning with contentful sentences uniquely activated a left hemisphere ventral network involving temporal (BA 21/22) and frontal regions (BA 44, 8, 9).⁷

⁷The initial encoding for content premise integration (content preparation condition) may also involve the left superior parietal lobe (BA 7). Examination of parameter estimates shows that the activation may be a function of “deactivation” of the parietal in the baseline condition involving the reading of unrelated sentences. However, since this was our low-level baseline, it is difficult to interpret the cause of the effect.

Examination of parameter estimates suggests that the activation of the striate cortex (BA 17) lingual gyri (BA 18) in the content minus no-content comparison reflects greater visual processing due to increased letter lengths in the content words. The parameter estimates also suggest that the activation in the other regions are related to reasoning and not word lengths (see Discussion of interaction below). Excluding the activation of the visual cortex, we are essentially left with the language system. BA 44 (Broca’s Area) is part of the ar-

Effect of Incongruency

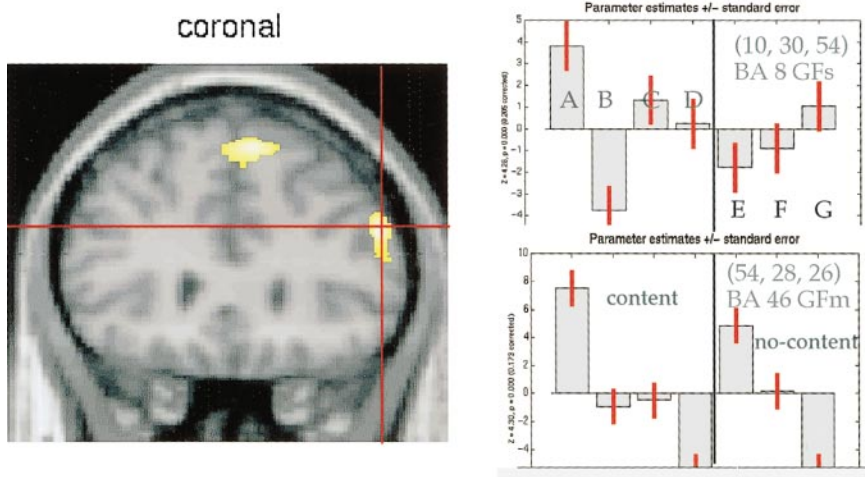


FIG. 6. SPMs rendered into standard stereotactic space and superimposed on to a coronal MRI in standard space. The direct comparison of the incongruent reasoning condition with the congruent reasoning condition (incongruent reasoning – congruent reasoning) reveals activation in right dorsal medial (10, 30, 54) and lateral prefrontal cortex (54, 28, 26). Parameter estimates show a differential response profile for the two regions. The dorsal medial prefrontal cortex is responsive to incongruent reasoning trials, while the right lateral prefrontal cortex is responsive to both incongruent and no-content reasoning trials. (A, Incongruent content reasoning; B, congruent content reasoning; C, content preparation; D, content baseline; E, no-content reasoning; F, no-content preparation; G, no-content baseline.)

tulatory loop of verbal WM (Paulesu *et al.*, 1993) and has also been associated with syntactic processing (Zurif, 1982). BA 21/22 (which overlaps with Wernicke's Area) is known to have semantic processing functions (Price, 1998; Schaffler *et al.*, 1996). This suggests that reasoning with contentful sentences is a linguistic process involving both syntactic and semantic processing. On its own, this finding is consistent with mental logic theory.

Two previous [¹⁵O] H₂O PET studies (Goel *et al.*, 1997, 1998) have reported similar frontal and temporal activations, but failed to show basal ganglia activation. (PET camera coverage limitations precluded them from reporting cerebellum activation.) However, another [¹⁵O] H₂O PET study (Osherson *et al.*, 1998), activated occipital, basal ganglia, and cerebellar regions, but failed to find activation of language areas (in the "logic vs meaning" condition). These differences may well be due to experimental design issues. For example, these studies used a block design in which the activation was averaged over 1 min, much of it being taken up by reading of the sentences, so it possible that activation associated with the reasoning component of the task was swamped by other aspects of the task.

When semantic content is lacking, a dorsal network incorporating bilateral occipital (BA 18, 19), left parietal (BA 7), bilateral dorsal frontal (BA 6), and left frontal (BA 44, 8, 10) and right frontal (BA 46) regions is used.⁸ This pattern of activation mirrors that involved in the internal representation and manipulation of spatial information (Jonides *et al.*, 1993; Kosslyn *et al.*, 1989; Laeng, 1994) and is very similar to that reported for certain types of mathematical reasoning involving approximation of numerical quantities (Dehaene *et al.*, 1999). This profile suggests that subjects construct a spatial representation from the sentences and is consistent with the claim that subjects solve abstract syllogisms through the use of Venn Diagrams, Euler Circles, or more generally, the spatial models predicted by mental model theory (Johnson-Laird, 1994).

It should be noted that the above network is also similar to patterns seen in some eye movement/visual attention tasks (Corbetta and Shulman, 1998), but such an interpretation is not consistent with the fact that content trials have greater visual eye movement requirements (due to longer sentence lengths) than no-content trials. Furthermore, the interaction analyses, which controls for the presence or absence of con-

tent words, and artifactual effects of differential letter lengths of content words and symbols in the content and no-content conditions respectively, also reveals language and spatial processing regions. This supports our interpretation that two distinct networks are involved in reasoning and that they are modulated by the presence and absence of content.

When the truth value or believability of the semantic content contradicted the validity judgement, subjects took longer to do the task and made more errors. This is consistent with other published behavioral data (Evans *et al.*, 1983). One explanation for the effect is that the content-based mechanism may be the default (Sloman, 1996). It will automatically engage unless there is an absence of content cues (as in the no-content trials) or when there is active suppression of content cues due to task demands (as in the incongruent trials). In either of these cases, the content mechanism has to be actively inhibited and the no-content mechanism recruited to complete the reasoning task.

Two regions of the right prefrontal cortex were activated by the conflict condition. Parameter estimates for these two regions reveal a very different response profile (see Fig. 6). Medial BA 8 responded positively only to the incongruent reasoning condition. In fact, it showed a negative response profile (compared to baseline) to the congruent reasoning condition, and was largely neutral in the no-content conditions. The lateral prefrontal cortex (BA 46/45) responded positively to both the incongruent reasoning and no-content reasoning conditions and indeed is part of the no-content reasoning network identified above. It was neutral with respect to the congruent reasoning condition. One possible interpretation of this differential response profile is that area BA 8 is part of an attention shifting or "executive function" mechanism involved in the detection of the belief-logic conflict, and the right inferior/middle prefrontal cortex (BA 46/45) is part of the reasoning mechanism recruited in the no-content condition. Another possibility is that BA 45/46 serves some conflict resolution function. Such a role of the right lateral prefrontal cortex has been noted in studies of maintenance of an intention in the face of conflict between action and sensory feedback (Fink *et al.*, 1999).

Our results predict that the left hemisphere is necessary and often sufficient for logical reasoning; the right hemisphere is sometimes necessary but never sufficient. There is remarkable consistency between our findings and lesion data. Read (1981) tested temporal lobectomy patients on three term relational problems with semantic content. He reported that LH patient performance was more impaired than RH patient performance. Gazzaniga and colleagues have administered simple reasoning tasks to split-brain patients and concluded that reasoning is a left hemisphere phenomenon (Gazzaniga, 1985; Gazzaniga and Smylie, 1984). For example, they report that the left hemi-

⁸ The initial encoding for no-content premise integration (no-content preparation condition) may also include the left superior parietal lobe (BA 7). But as above, examination of parameter estimates for the condition shows that the activation is a function of "deactivation" of the parietal in the low-level baseline condition involving the reading of unrelated sentences, therefore making interpretation of the effect difficult.

sphere will readily infer “boiling water” when presented with “water” and “pan,” while the right hemisphere seems incapable of such inferences. Gazzaniga goes on to postulate a “left brain interpreter,” a mechanism that continuously elaborates and interprets information presented to it and readily draws inferences. Caramazza *et al.* (1976) administered two-term problems such as the following: “Mike is taller than George”; who is taller? They reported that left hemisphere patients were impaired in all forms of the problem but right hemisphere patients were only impaired when the form of the question was incongruent with the premise (e.g., who is shorter?). This is precisely what our results would predict.

In conclusion, mental logic theories predict that the language (syntactic) system is both necessary and sufficient for deductive reasoning while mental model theories predict that the visuo-spatial system is necessary and sufficient. Contrary to these theories, but consistent with the dual mechanism theory, we have found evidence for the engagement of both systems. The presence of semantic content engages the language system in the reasoning process. The absence of semantic content engages the spatial system in the identical reasoning task. This dissociation is remarkable because the logically relevant information is identical in both conditions. These mechanisms may be differentially sensitive to reasoning about either (i) abstract vs concrete contents or (ii) contents that one has beliefs about vs nonreferential contents, which cannot give rise to beliefs. While our design does not allow us to differentiate between these two possibilities, the fact that we find a neural basis for belief–logic conflicts suggests that the latter may be the correct interpretation.

ACKNOWLEDGMENTS

We are indebted to Richard Frackowiak for internal review of this manuscript. VG was supported in part by a York University Faculty of Arts Fellowship and NSERC and SSHRC grants. CB, CF, & RJD are supported by the Wellcome Trust.

REFERENCES

- Ashburner, J., and Friston, K. J. 1999. Nonlinear spatial normalization using basis functions. *Hum. Brain Mapp.* 7(4): 254–266.
- Caramazza, A., Gordon, J., Zurif, E. B., and DeLuca, D. 1976. Right-hemispheric damage and verbal problem solving behavior. *Brain Lang.* 3(1): 41–46.
- Channon, S., Jones, M. C., and Stephenson, S. 1993. Cognitive strategies and hypothesis testing during discrimination learning in Parkinson's disease. *Neuropsychologia* 31(1): 75–82.
- Corbetta, M., and Shulman, G. L. 1998. Human cortical mechanisms of visual attention during orienting and search. *Philos. Trans. R. Soc. London B Biol. Sci.* 353(1373): 1353–1362.
- Cummings, J. L. 1993. Frontal-subcortical circuits and human behavior. *Arch. Neurol.* 50(8): 873–880.
- Dehaene, S., Spelke, E., Pineda, P., Stanescu, R., and Tsivkin, S. 1999. Sources Of Mathematical Thinking: Behavioral And Brain-Imaging Evidence. *Science* 284(May 7): 970–974.
- Dickstein, L. S. 1978. The Effect of Figure on Syllogistic Reasoning. *Memory Cognition* 6: 76–83.
- Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., and Peters, T. M. 1993. 3D Statistical Neuroanatomical Models from 305 MRI Volumes. In *Proc. IEEE-Nuclear Science Symposium and Medical Imaging Conference*, pp. 1813–1817.
- Evans, J. S. B. T., Barston, J., and Pollard, P. 1983. On the conflict between logic and belief in syllogistic reasoning. *Memory Cognition* 11: 295–306.
- Fiez, J. A., Petersen, S. E., Cheney, M. K., and Raichle, M. E. 1992. Impaired nonmotor learning and error detection associated with cerebellar damage. A single case study. *Brain* 115(Pt. 1): 155–178.
- Fink, G. R., Marshall, J. C., Halligan, P. W., Frith, C. D., Driver, J., Frackowiak, R. S., and Dolan, R. J. 1999. The neural consequences of conflict between intention and the senses. *Brain* 122(Pt. 3): 497–512.
- Friston, K., Holmes, A., Worsley, K., Poline, J.-B., Frith, C., and Frackowiak, R. 1995. Statistical parametric maps in functional imaging: A general approach. *Hum. Brain Mapp.* 2: 189–210.
- Gazzaniga, M. S. 1985. *The Social Brain*. Basic Books, New York.
- Gazzaniga, M. S., and Smylie, C. S. 1984. Dissociation of language and cognition. *Brain* 107: 145–153.
- Goel, V. 1995. *Sketches of Thought*. MIT Press, Cambridge, MA.
- Goel, V., Gold, B., Kapur, S., and Houle, S. 1997. The seats of reason: A localization study of deductive & inductive reasoning using PET (O15) blood flow technique. *NeuroReport* 8(5): 1305–1310.
- Goel, V., Gold, B., Kapur, S., and Houle, S. 1998. Neuroanatomical correlates of human reasoning. *J. Cogn. Neurosci.* 10(3): 293–302.
- Grafman, J., Litvan, I., Massaquoi, S., Stewart, M., Sirigu, A., and Hallett, M. 1992. Cognitive planning deficit in patients with cerebellar atrophy [see Comments]. *Neurology* 42(8): 1493–1496.
- Johnson-Laird, P. N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard Univ. Press, Cambridge, MA.
- Johnson-Laird, P. N. 1994. Mental Models, Deductive Reasoning, and the Brain. In *The Cognitive Neurosciences* (M. S. Gazzaniga, Eds.), pp. 999–1008. MIT Press, Cambridge, MA.
- Johnson-Laird, P. N., and Byrne, R. M. J. 1991. *Deduction*. Lawrence Erlbaum, Hillsdale, NJ.
- Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., and Minoshima, S. 1993. Spatial working memory in humans as revealed by PET. *Nature* 363: 623–625.
- Kosslyn, S. M., Koenig, O., Cave, C. B., Tang, J., and Gabrieli, J. D. E. 1989. Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *J. Exp. Psychol. Hum. Percept. Performance* 15(4): 723–735.
- Laeng, B. 1994. Lateralization of categorical & coordinate spatial functions: A study of unilateral stroke patients. *J. Cogn. Neurosci.* 6(3): 189–203.
- Middleton, F. A., and Strick, P. L. 1994. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science* 266(5184): 458–461.
- Newell, A. 1980. Reasoning, problem solving, and decision processes: The problem space as a fundamental category. In *Attention and Performance VIII* (R. S. Nickerson, Eds.), Lawrence Erlbaum, Hillsdale, NJ.
- Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., and Fazio, F. 1998. Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia* 36(4): 369–376.
- Paulesu, E., Frith, C. D., and Frackowiak, R. S. J. 1993. The neural correlates of the verbal component of working memory. *Nature* 362: 342–345.
- Price, C. J. 1998. The functional anatomy of word comprehension and production. *Trends Cogn. Sci.* 2(8): 281–288.

- Rao, S. M., Bobholz, J. A., Hammeke, T. A., Rosen, A. C., Woodley, S. J., Cunningham, J. M., Cox, R. W., Stein, E. A., and Binder, J. R. 1997. Functional MRI evidence for subcortical participation in conceptual reasoning skills. *Neuroreport* **8**(8): 1987–1993.
- Read, D. E. 1981. Solving deductive-reasoning problems after unilateral temporal lobectomy. *Brain Language* **12**: 116–127.
- Sandson, T. A., Daffner, K. R., Carvalho, P. A., and Mesulam, M. M. 1991. Frontal lobe dysfunction following infarction of the left-sided medial thalamus. *Arch. Neurol.* **48**(12): 1300–1303.
- Schaffler, L., Luders, H. O., and Beck, G. J. 1996. Quantitative comparison of language deficits produced by extraoperative electrical stimulation of Broca's, Wernicke's, and basal temporal language areas. *Epilepsia* **37**(5): 463–475.
- Sloman, S. A. 1996. The empirical case for two systems of reasoning. *Psychol. Bull.* **119**(1): 3–22.
- Worsley, K. J., and Friston, K. J. 1995. Analysis of fMRI time-series revisited—Again. *NeuroImage* **2**: 173–181.
- Zurif, E. B. 1982. Language and brain: Some points of connection. In *Perspectives on Mental Representation: Experimental and Theoretical Studies of Cognitive Processes and Capacities* (J. Mehler, E. C. T. Walker, and M. Garrett, Eds.), Lawrence Erlbaum, Hillsdale, NJ.