Binding crossmodal object features in perirhinal cortex

Kirsten I. Taylor*†‡, Helen E. Moss*, Emmanuel A. Stamatakis*§, and Lorraine K. Tyler*§

*Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom; †Memory Clinic–Neuropsychology Center, University Hospital Basel, Schanzenstrasse 55, CH-4031 Basel, Switzerland; and §Wolfson Brain Imaging Centre, University of Cambridge, Cambridge CB2 2QQ, United Kingdom

Knowledge of objects in the world is stored in our brains as rich, multimodal representations. Because the neural pathways that process this diverse sensory information are largely anatomically distinct, a fundamental challenge to cognitive neuroscience is to explain how the brain binds the different sensory features that comprise an object to form meaningful, multimodal object representations. Studies with nonhuman primates suggest that a structure at the culmination of the object recognition system (the perirhinal cortex) performs this critical function. In contrast, human neuroimaging studies implicate the posterior superior temporal sulcus (pSTS). The results of the functional MRI study reported here resolve this apparent discrepancy by demonstrating that both pSTS and the perirhinal cortex contribute to crossmodal binding in humans, but in different ways. Significantly, only perirhinal cortex activity is modulated by meaning variables (e.g., semantic congruency and semantic category), suggesting that these two regions play complementary functional roles, with pSTS acting as a presemantic, heteromodal region for crossmodal perceptual features, and perirhinal cortex integrating these features into higher-level conceptual representations. This interpretation is supported by the results of our behavioral study: Patients with lesions, including the perirhinal cortex, but not patients with damage restricted to frontal cortex, were impaired on the same crossmodal integration task, and their performance was significantly influenced by the same semantic factors, mirroring the functional MRI findings. These results integrate nonhuman and human primate research by providing converging evidence that human perirhinal cortex is also critically involved in processing meaningful aspects of multimodal object representations.

conceptual knowledge | hierarchical object processing | ventral stream

A major outstanding question in the cognitive neurosciences is how different unimodal object features are integrated into coherent, multimodal object representations. Hierarchical models of object processing based on studies with nonhuman primates suggest that the perirhinal cortex, located at the culmination of the ventral occipitotemporal object-processing stream, performs this critical function. Within this stream, increasingly more complex combinations of visual object features are processed from posterior to anterior ventral temporal lobe sites (1–3), with perirhinal cortex of the anteromedial temporal lobe integrating the most complex combinations of features required for fine-grained visual discriminations between objects (4, 5). Recent functional MRI (fMRI) and lesion studies generally support this model in the human system. Lerner *et al.* (6) demonstrated that the sensitivity of ventral occipito-temporal regions to the scrambling of car images increased significantly from posterior (V1, V2, V3, V4/V8) to more anteriorly situated sites (lateral occipital sulcus and posterior fusiform gyrus; lateral occipital complex), with scrambled images predicting activity in more posterior sites and intact images predicting activity in the more anterior regions. The hypothesized role of anteromedial structures in complex visual discriminations was confirmed in another series of fMRI experiments

and in neuropsychological studies with brain-damaged patients (7, 8). In the fMRI studies, tasks that did not require complex feature conjunctions (e.g., distinguishing living from nonliving things, which can be accomplished on the basis of general featural differences, such as curvature) only activated posterior temporal and occipital regions, whereas tasks that required complex conjunctions of features (e.g., the combination of features necessary to distinguish between highly similar objects, such as a lion and a tiger) additionally activated the anteromedial temporal lobe, including the perirhinal cortex. These findings were confirmed in behavioral experiments with patients with lesions including the perirhinal cortex, who were unable to perform complex visual discriminations (e.g., distinguishing between a lion and a tiger) while retaining the ability to perform simple visual discriminations (e.g., those necessary to distinguish between living and nonliving things; refs. 7 and 8).

Motivated by the multimodal nature of our conceptual representations, it has been proposed that similar hierarchical processing pathways operate, not only in the human visual processing stream, but also in each sensory modality (9, 10). For example, nonhuman primate work has demonstrated that, within the auditory modality, a hierarchical system progressing from core and belt through lateral parabelt regions to the anterior superior temporal gyrus is likewise involved in identifying increasingly more complex auditory stimuli ranging from pure to complex tones to species-specific communication calls (11–13). In humans, this auditory stream appears to progress in a similar fashion from the primary auditory cortex in the transverse temporal (Heschl's) gyrus to the surrounding secondary auditory cortex in the superior temporal gyrus (14), but then diverges from the nonhuman primate system by sending information primarily posteriorially into the superior temporal gyrus for the identification of spoken language and environmental sounds (15 and 16, respectively; 12). Hierarchical object-processing models (e.g., 10) propose that multimodal representations are formed at convergence sites where information from each of the sensory streams is integrated. Critically, a key site in the ventral stream that receives inputs from all other sensory modalities via unimodal or polymodal association areas is the perirhinal cortex (17), suggesting that this structure may also integrate information across sensory modalities to form multimodal object representations (5). Ablation studies in nonhuman primates support this claim: Monkeys with bilaterally aspirated anterior rhinal cortices (perirhinal and entorhinal cortices) were severely impaired in relearning a crossmodal tactile-visual delayed nonmatching to sample task compared with intact and bilaterally

NEUROSCIENCE

amygdalectomized control animals (18; see also ref. 19). Moreover, an earlier study by Desimone and Gross (20) demonstrated that single perirhinal cortex neurons evidenced multimodal properties, responding to both visual and auditory stimuli.

In contrast, human functional imaging studies have identified a different region associated with the crossmodal integration of audiovisual object features: the posterior superior temporal sulcus (pSTS) extending into the middle temporal gyrus (MTG). This region responds more strongly to combinations of sounds and pictures or videos of objects compared with unimodal presentations of the same stimuli (21, 22), mimicking the behavior of some single neurons in the polysensory area in the nonhuman primate STS (23). However, the pSTS/MTG appears to be relatively insensitive to the meaning of multimodal objects, because activity here is neither significantly modulated by the semantic relationship between crossmodal stimuli (e.g., a video of a saw and a sawing sound vs. a video of a hammer and a sawing sound; ref. 21), nor the semantic category to which the crossmodal stimuli belong (21, 22). Because the ultimate goal in the crossmodal integration of object features must be the creation of meaningful, multimodal object representations, it remains unclear whether the pSTS/MTG plays a role in crossmodal integration beyond that of a heteromodal (polymodal) region, i.e., one receiving inputs from more than one sensory modality without synthesizing these into novel, multimodal representations. Indeed, lesions of the STS heteromodal region in animals have not produced unequivocal impairments in crossmodal integration abilities (18, 24).

The key, and as yet unaddressed, question that we ask in this study is whether the human perirhinal cortex functions as a crossmodal integration site for the binding of perceptual object properties into conceptual representations. In an event-related fMRI (efMRI) study with healthy participants and a behavioral study with patients, we asked two related questions: (*i*) Does human perirhinal cortex participate in the crossmodal integration of object features, as suggested by the nonhuman primate literature? (*ii*) Is the extent of perirhinal cortex involvement in crossmodal integration affected by the meaning of the objects? We would expect the latter to be true if the ultimate goal in the crossmodal integration of object features is the creation of meaningful (semantic), multimodal object representations. Our efMRI study presented 15 healthy participants with pairs of sounds and pictures (e.g., the sound "roar" and a picture of a lion) in the crossmodal conditions, and two parts of a sound and two parts of a picture in the unimodal baseline conditions. Half of the stimuli were congruent (e.g., the sound "meow" and a picture of a cat), and half of the stimuli were incongruent (e.g., the sound "woof" and a picture of an elephant). Participants decided, for every trial, whether the two items were congruent or incongruent (i.e., whether they "go together") by pressing different response keys, thus ensuring that participants attended to and attempted to integrate the two stimuli in each condition. We measured neural activity during the crossmodal integration conditions and compared it with that during the unimodal integration conditions to identify sites specifically involved in integrating object features across modalities while controlling for integration *per se* and associated decision-making processes. The congruency manipulation served a further critical function: By measuring responses to stimuli that were either meaningfully related (congruent) or not meaningfully related (incongruent), we could evaluate the responsiveness of different sites to the semantic relationship between the crossmodal stimuli. A further semantic manipulation was introduced by varying the semantic category (living and nonliving) of the objects. Recent research has shown that living things (especially animals) typically have many more features than nonliving things, and that living things have many more shared properties (e.g., many animals have four legs, eyes, and the ability to hunt) compared with nonliving things (25, 26). The higher degree of featural overlap increases the similarity between living things and makes them relatively more

difficult to distinguish from one another. In light of these findings, we predicted that the visual and auditory stimuli representing living things would activate much larger clusters of associated semantic features than the crossmodal stimuli representing nonliving things. Moreover, because living things share more features with one another than nonliving things, the activated features themselves will be more ambiguous. Because these crossmodal sets of features must be integrated with one another to perform the task, we predicted that the greater size and ambiguity of the sets of semantic features associated with living things would tax the crossmodal integration processes of the perirhinal cortex (18) to a greater extent than those of nonliving things, leading to increased perirhinal cortex activity. The unimodal visual baseline task ensured that the perirhinal cortex activity could be attributed to the complexity of crossmodal, and not only intramodal, visual feature integration processes (8).

To further investigate the putative role of the perirhinal cortex in crossmodal integration, we presented two herpes simplex encephalitis (HSE) patients (with anteromedial lesions including the perirhinal cortex) and two patients [with spared perirhinal cortices and lesions in the left inferior frontal cortex (LIFC)] with shortened versions of the same crossmodal and unimodal tasks, and compared their performances with those of 12 mature control participants. If the perirhinal cortex is critically involved in the crossmodal integration of meaningful object features, then only the HSE patients should be impaired on this task, and the magnitude of their deficit should interact with the semantic factors of congruency and category.

Results

efMRI Study. The behavioral data (accuracy and reaction times) are shown in Table 1, which is published as supporting information on the PNAS web site. Critically, there were no significant differences in error rates and reaction times in the crossmodal integration conditions for congruent and incongruent stimuli [91.5% vs. 92.9% correct, respectively, t (14) = 0.80, P = not significant (ns); and 898 ms vs. 920 ms, respectively, t (14) = 1.43, P = ns] and for living and nonliving stimuli [92.1% vs. 92.3% correct, respectively, t (14) = 0.24, P = ns; and 898 ms vs. 920 ms, respectively, t (14) = 1.37, P = ns].

The random effects analysis comparing crossmodal integration with unimodal auditory and visual integration resulted in three significant clusters of activation. One cluster encompassed the bilateral medial frontal lobes and anterior cingulate gyri (BA 10/32; peak voxel at 0, 54, 2), a region previously ascribed a monitoring function (27, 28). This functional-neuroanatomical relationship is supported by the finding that activation in this region was driven primarily by the incongruent crossmodal condition: Whereas incongruent crossmodal compared with unimodal integration resulted in a comparable cluster of medial frontal and anterior cingulate activity (peak voxel at −2, 58, 0), the contrast of congruent crossmodal compared with congruent unimodal integration did not yield suprathreshold activity in this region. A second cluster was centered in the left posterior MTG and included the lower bank of the pSTS extending posteriorly into the peristriate cortex (BA 39, 19; peak voxel at −46, −76, 22; Fig. 1). To determine whether this region was responsive to the semantic factors of congruency and category, effect sizes in the corresponding conditions were calculated with MARSBAR[¶] and compared. These effect sizes show that activation in this region was sensitive neither to the congruency of the crossmodal stimulus pair (congruent vs. incongruent: t (14) = 0.65; P = 0.26) nor to the living/nonliving manipulation (living vs. nonliving: t (14) = 0.37, P = 0.36; Fig. 1 *a* and *b*). This pattern of results demonstrates that the pSTS/MTG is insensitive to semantic factors, i.e., whether the crossmodal stimuli are meaningfully

---

[¶]Brett, M., Anton, J.-L., Valabregue, R. & Poline, J.-B. (2002) *NeuroImage* **16**, S497 (abstr.).
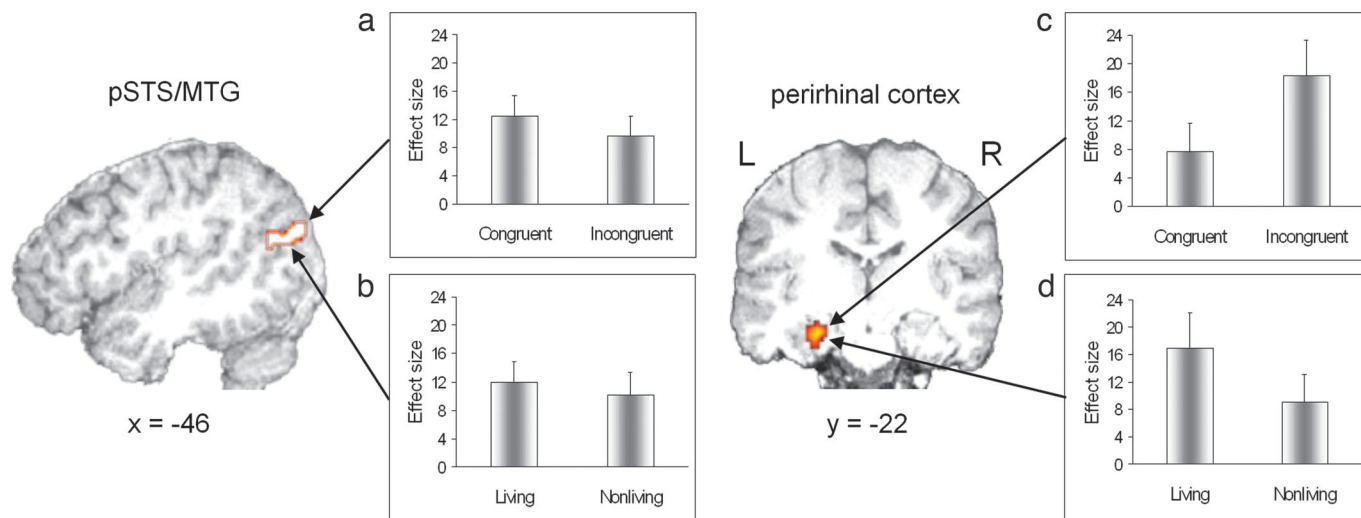
Fig. 1. Responsiveness to meaning during crossmodal integration. (*Left*) Responses in the left pSTS/MTG were insensitive to whether crossmodal pairs were meaningfully related (i.e., congruency) (*a*) and object category (i.e., living/nonliving) (*b*). (*Right*) Activity centered in the left perirhinal cortex (BA 35, 36) (activation thresholded at an uncorrected voxel $P < 0.01$ for display purposes) signaled whether the crossmodal inputs were meaningfully related (*c*) and was greater for living than nonliving things (*d*). Clusters are rendered on a single participant's mean-normalized anatomic image. Error bars represent standard errors. Montreal Neurological Institute coordinates are reported.

related and the conceptual category to which they belong. We also calculated the interaction contrast suggested by Calvert (29), which confirmed the significant activation in this cluster ($t = 4.93, P = 0.0001$).

A significant cluster of activation was also located in the left perirhinal cortex at a peak voxel of $-26, -20, -22$ (uncorrected voxel-level $P < 0.001$). Given our *a priori* hypothesis for activity in this region, we calculated a small volume correction with a 10-mm sphere centered at the peak voxel of this cluster (30, 31). This analysis revealed that activity in this region was significant at a corrected cluster-level of $P < 0.05$. The interaction contrast suggested by Calvert (29) also demonstrated significant activity in this region ($t = 3.34, P = 0.002$). Finally, closer examination of the fixed effects analyses (at the individual subject level) confirmed perirhinal activity in single participants (see Fig. 3, which is published as supporting information on the PNAS web site).

We next investigated the responsiveness of perirhinal cortex activity to the semantic variables of congruency and category. Again, these analyses were based on estimates of effect sizes in the perirhinal cortex during the relevant crossmodal conditions as calculated with MARSBAR[‖]. These analyses revealed that, in contrast to activation in the pSTS/MTG, activity in the perirhinal cortex was modulated by the meaningfulness of the stimuli: Living things evoked stronger responses than nonliving things [$t (14) = 1.83; P = 0.04$], and there was a trend toward stronger activation for responses to incongruent compared with congruent stimuli [$t (14) = 1.29, P = 0.11$; see Fig. 1 *c* and *d*]. These semantic effects, which were not seen in the pSTS/MTG, suggest that crossmodal integration in the perirhinal cortex is modulated by the semantic content of the crossmodal stimuli, with greater responses to objects with more and more similar, overlapping features (i.e., living things) whose identification requires more complex conjunctions of crossmodal features, as well as the semantic relationship (i.e., congruency) between the crossmodal stimuli.

Behavioral Study. The mature control (MC) group's and patients' percentage of correct performances are shown in Table 2, which is published as supporting information on the PNAS web site. The multivariate analyses of the MC patients' accuracy perfor-

mances revealed a significant main effect of task (crossmodal vs. unimodal) in the participant means ($F_1 = 5.74, P < 0.05$), which was not significant in the items' analysis ($F_2 = 0.83, P = 0.36$). The main effects of congruency and category, as well as all interactions, were not significant.

The HSE and LIFC patients' performances were compared with normal performance by converting their scores to z scores (standard scores), which represent the magnitude of the difference between patient performance and the MC participants' performances in terms of control participant standard deviations {e.g., [(mean performance of control participants) − (mean performance of patients)]/(standard deviation of control participants' performance)}. HSE patients' z scores indicated that they were impaired on both the unimodal ($z = -2.85; P < 0.01$) and crossmodal ($z = -9.39; P \ll 0.0001$) integration tasks. Within the unimodal tasks, their performances were impaired on both incongruent ($z = -2.25, P < 0.05$) and nonliving trials ($z = -3.98, P < 0.0001$). These differences appeared to be because of the HSE patients' particularly poor performance on the incongruent, living trials of the unimodal auditory baseline with words ($z = -7.28; P < 0.0001$), as well as the nonliving trials of the unimodal visual baseline task ($z = -2.22, P < 0.05$ and $z = -3.37, P < 0.001$ for congruent and incongruent trials, respectively). HSE patients' performances were impaired on all crossmodal integration conditions (all $z < -3.50$; all $P < 0.001$). In contrast, LIFC patients' performances on all tasks and conditions did not significantly differ from those of the MC participants' (all $P =$ ns).

The results of the χ^2 analyses demonstrated that the HSE patients were more impaired on the crossmodal compared with the unimodal tasks ($\chi^2 = 8.99; P < 0.01$). Within the unimodal integration tasks, HSE patients performed comparably on congruent compared with incongruent trials ($\chi^2 = 1.15; P =$ ns) and on living compared with nonliving trials ($\chi^2 = 1.15; P =$ ns). However, within the crossmodal tasks, HSE patients performed significantly more poorly on the incongruent compared with congruent trials ($\chi^2 = 8.04; P < 0.01$). This effect does not appear to be due to a "congruent" response bias, because there was no congruency effect in the unimodal trials. There was a trend for the HSE patients to perform more poorly on living compared with nonliving crossmodal trials (65.0% vs. 77.5%, respectively, $\chi^2 = 2.63; P = 0.11$). In
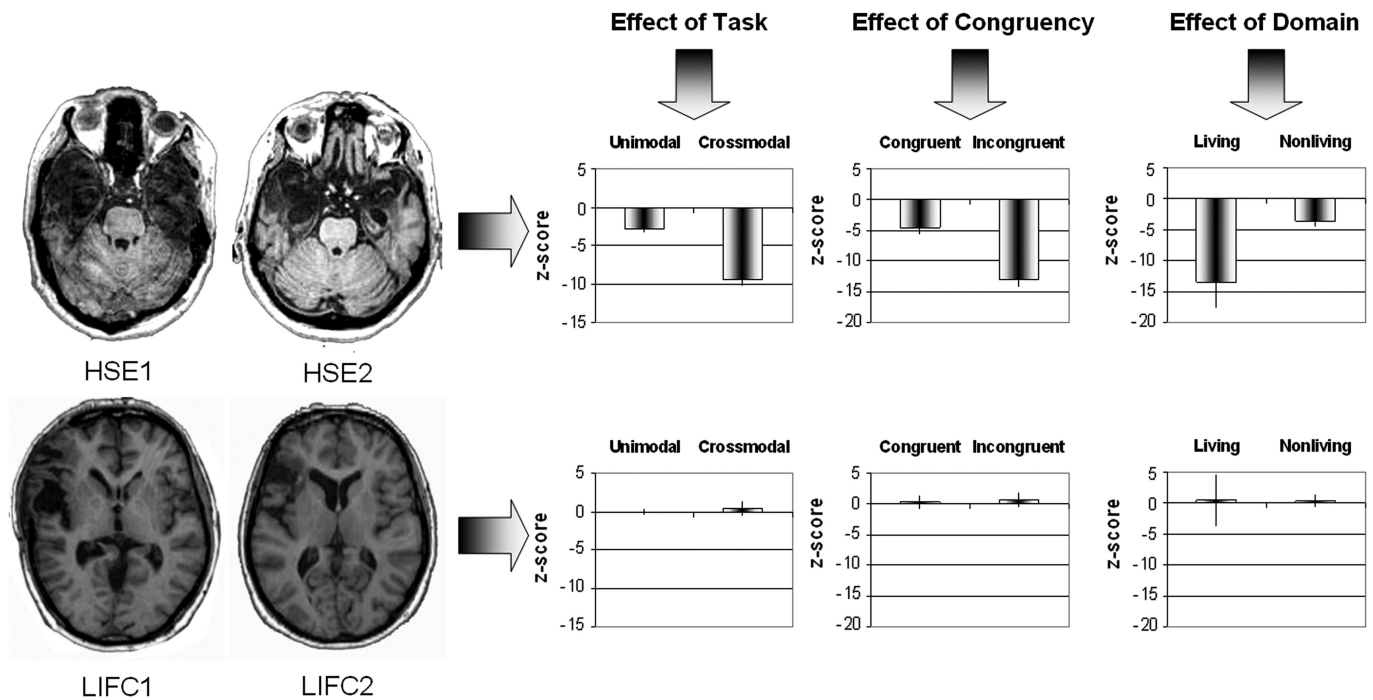
Fig. 2. The crossmodal integration performances of patients with perirhinal cortex lesions mirrored the efMRI findings. Two herpes simplex encephalitis patients (HSE1, HSE2) with lesions including the perirhinal cortex were disproportionately impaired on crossmodal compared with unimodal integration tasks and, on the crossmodal tasks, performed worse on incongruent than congruent trials and with living compared with nonliving stimulus pairs (effects of task, congruency, and category, respectively). Two patients with lesions centered in the left inferior frontal cortex (LIFC1, LIFC2) serving as positive control participants performed comparably on both integration tasks and in both congruency and category conditions. Patients' performances are quantified in terms of the control participant performances (z scores ± SD).

contrast, MC participants performed more consistently on the living compared with the nonliving crossmodal trials (standard deviations of 2.3 and 5.0, respectively; see Table 2), suggesting that smaller decrements in performance on living trials are required in order for them to be considered "impaired." This interpretation is reflected in the HSE patients' z scores, which indicated much greater impairments on the living compared with the nonliving crossmodal trials ($z = -13.46, P < 1.0 \times e^{-41}$ for living crossmodal trials and $z = -3.56, P < 0.001$ for the nonliving crossmodal trials). LIFC patients, on the other hand, performed comparably on the unimodal and crossmodal tasks ($\chi^2 = 0.86; P = $ ns) and, within each integration task, on the congruent and incongruent trials and on the living and nonliving trials (all $\chi^2 < 1.03$; all $P = $ ns). Taken together, the effects of task, congruency, and category evident in the HSE patients' performances mirror the efMRI findings (see Fig. 2).

Discussion

The present results extend the nonhuman data on the functional-neuroanatomical basis of crossmodal integration of object features into the human domain. Our efMRI study with healthy participants demonstrated perirhinal cortex activity, confirmed in single subjects, during crossmodal compared with unimodal integration of audiovisual object features. The importance of perirhinal cortex involvement in crossmodal integration was confirmed by the behavioral performances of HSE patients with lesions including this region, who were more impaired during crossmodal compared with unimodal integration tasks. Together, these findings provide converging evidence that, like nonhuman primates, the human perirhinal cortex is critically involved in binding the auditory and visual features of real objects together.

The results of the current efMRI experiment revealed that a distributed temporal lobe network, including the pSTS/MTG

and perirhinal cortex, supports crossmodal integration and suggests that these regions play complementary functional roles. pSTS/MTG activity during crossmodal integration was not modulated by semantic congruency (i.e., whether the crossmodal stimuli were meaningfully related or not) or category (living vs. nonliving object features). This pattern is consistent with recent reports that also failed to find significant effects of congruency or stimulus category (i.e., videos of tools vs. faces or of moving tools vs. human bodies) in this region (refs. 22 and 21, respectively). The lack of congruency and semantic category effects in the present and previous (21, 22) studies suggest that, during the crossmodal integration of object features, the pSTS/MTG functions as a presemantic, heteromodal sensory area.

Patterns of activity in the perirhinal cortex were strikingly different from those within pSTS/MTG. Perirhinal responses were sensitive to the meaningfulness of the crossmodal stimuli: There was a trend for responses in the unrelated (incongruent) crossmodal condition to be stronger than in the meaningfully related (congruent) crossmodal condition, suggesting a sensitivity to the semantic relationship between the visual and auditory stimuli. Importantly, activity in the perirhinal cortex was also greater during the crossmodal integration of living compared with nonliving things. A number of studies have shown that living things have many more features than nonliving things and that living things share more features with one another than nonliving things (7, 8). Some of these features refer to visual features (e.g., has eyes, has tail), whereas others refer to nonvisual properties (e.g., growls, hunts). Recent fMRI studies have demonstrated greater perirhinal cortex activation during the visual processing of living compared with nonliving things, consistent with a role of this structure in complex visual integration processes necessary to discriminate and identify more visually complex living objects (7, 8). However, the category effect in the crossmodal

condition is unlikely to be due to complex visual integration processes, because the unimodal visual baseline task controlled for these processes. In fact, healthy subjects ($n = 20$) actually rated the unimodal visual baseline stimuli ($n = 100$) as slightly more visually complex than 100 whole pictures from the crossmodal conditions [$F_1 (1, 19) = 9.07, P = 0.007; F_2 (1, 196) = 8.36, P = 0.004$], and, although these visual complexity ratings interacted with the living/nonliving category in the subjects' analyses [$F_1 (1, 19) = 18.37, P < 0.001$], this effect was because of the difference between visual complexity ratings for picture halves vs. wholes being greater for nonliving compared with living things. Moreover, an additional fMRI model with visual complexity of the whole pictures as a parametric modulator revealed that perirhinal activity was not modulated by the visual complexity of pictures of living and nonliving things, even when small volume corrections were applied. Instead, we suggest that the crossmodal integration condition with living things activated much larger and more ambiguous sets of semantic features related to the living visual and auditory stimuli compared with the crossmodal condition with nonliving things, thereby placing greater demands on the crossmodal integration processes supported by the perirhinal cortex. These semantic effects were confirmed in our behavioral study in which patients with perirhinal cortex lesions performed worse on crossmodal compared with unimodal tasks and, within the crossmodal tasks, performed worse with incongruent than congruent and with living than nonliving crossmodal trials. Thus, these behavioral findings mirror the patterns of perirhinal cortex activity in the fMRI study. Taken together, the findings from these two studies provide converging evidence that human perirhinal cortex plays a critical role in binding the meaningful aspects of audiovisual object features together to form coherent, multimodal object representations.

The single-unit recording study by Higuchi and Miyashita (32) illustrates how crossmodal binding in perirhinal cortex might be achieved. Anterior commissurotomized monkeys were taught to associate different pairs of visual fractal pattern stimuli (e.g., S–S′). After this learning phase, some neurons in the inferotemporal lobe showed "pair-coding" properties (33); that is, they responded as strongly to S as to S′ when these stimuli were presented in isolation. The rhinal sulci were then unilaterally lesioned, animals relearned the old stimulus set and a new stimulus set, and neurons were again recorded from the same region of the inferotemporal lobe. Remarkably, although these neurons responded normally to individual visual stimuli postoperatively, they no longer exhibited pair-coding properties, neither for the preoperatively nor the postoperatively learned stimulus pairs. These findings strongly suggest that the rhinal cortex is responsible for binding visual stimuli together, presumably via backward neural signals to visual representations coded in posterior sites. The current results suggest that the role of the perirhinal cortex may extend beyond one of maintaining purely intramodal visual associations to that of a "master binder," integrating not only visual but also the polymodal inputs it receives (17) into multimodal stimulus associations necessary to represent objects in semantic memory (4, 5).

The present results are in line with the conceptualization of a division of labor within the anteromedial temporal lobe, with perirhinal cortex supporting semantic object memories and with more downstream structures, i.e., the entorhinal cortex and hippocampal structures, supporting episodic memory (but see e.g., ref. 34 for an alternative account). Several lines of evidence from the animal connectivity literature also support this distinction. Firstly, the macaque perirhinal cortex receives the majority of inputs from unimodal sensory regions representing unisensory object features, i.e., the anterior ventral temporal lobe (visual information), the superior temporal gyrus (auditory information), and the insular cortex (somatosensory information), and a smaller number of inputs from polymodal association regions (orbitofrontal cortex, dorsal STS, cingulate cortex, and the parahippocampal cortex). The entorhinal cortex, on the other hand, receives the majority of its inputs from higher-order, heteromodal association areas (orbitofrontal, parainsular, cingulate, retrosplenial, perirhinal, and parahippocampal cortices and the dorsal STS), with only unimodal olfactory information being sent directly here (olfactory bulb and piriform cortex). This information is then relayed along the perforant path to the dentate gyrus, hippocampus, and subiculum. Secondly, perirhinal and entorhinal cortices and the hippocampal structures can be characterized by a network of intrinsic associative connections, indicating that the information each receives is integrated in the respective structure. Thirdly, the perirhinal cortex sends information to the entorhinal cortex (35), and the entorhinal cortex sends information to the hippocampal structures, via feedforward (i.e., ascending) projections characteristic of hierarchical processing systems. Lavanex and Amaral (36) used these characteristics to conceptualize this system as a "hierarchy of connectivity." Taken together, we suggest that these connectivity findings indicate that information reaching the perirhinal cortex, i.e., primarily unimodal object feature information, is both necessary and sufficient for the semantic representation of objects, i.e., semantic memories of objects. Afferents to the entorhinal cortex, i.e., inputs from the perirhinal cortex as well as other, higher-order association areas, suggest that the entorhinal cortex and downstream hippocampal structures bind other, associative or contextual information together with the semantic object representation, providing both necessary and sufficient conditions for the formation of episodic memories. Thus, the afferent, intrinsic and efferent connectivity of the perirhinal cortex strongly suggests that it is primarily responsible for processing object-related featural information both necessary and sufficient to represent these objects in memory.

The multimodality of our semantic memories for objects undoubtedly conferred an evolutionary advantage, e.g., fleeing when we either heard a roar or saw a tiger, irrespective of the context in which these stimuli were encountered. The present results extend findings from the nonhuman primate literature into the human domain by demonstrating that the human perirhinal cortex is also critically involved in the crossmodal integration of object features. This function of the perirhinal cortex and its sensitivity to semantic variables shown here, together with the connectivity findings reviewed in the *Discussion*, suggest that it integrates perceptual feature information into higher-level, semantic memories of meaningful objects (refs.18 and 37; see also ref. 38).

Methods

efMRI Study. *Participants.* Fifteen right-handed, healthy participants (aged between 18 and 31 years; 5 males) participated. All gave informed consent. The study was approved by Addenbrooke's National Health Service (Cambridge, U.K.) Trust Ethical Committee.

Materials. The crossmodal stimuli consisted of 200 color photographs of objects, each paired with a specific property, half of which were environmental sounds and half of which were spoken words. Durations of environmental sounds and spoken words were matched. Stimuli in the unimodal visual ($n = 100$) and unimodal auditory ($n = 100$) conditions were constructed by halving stimuli from the respective modality in the crossmodal conditions (i.e., auditory and visual), and presenting two stimulus halves for congruency decisions. All 200 crossmodal and all 200 unimodal stimuli were unique. Half the trials in each condition were congruent (e.g., the sound "moo" and picture of a cow) and half were incongruent. Within each congruency condition, half of the stimuli represented living things and half represented nonliving things. We also included simple baseline stimuli consisting of arrows pointing in the same ($n = 50$) or different ($n = 50$) directions as well as 50 trials of rest.

Procedure. Auditory and visual stimuli were presented simultaneously in the crossmodal conditions. In the unimodal visual condition, the two stimulus halves were presented simultaneously, and in the auditory condition, the two halves were presented sequentially, separated by 750 ms of silence. Picture stimuli were displayed for 1,800 ms, and the mean duration of the unimodal auditory baseline was 1,445 ms. We used a sparse imaging design to avoid the confounding effects of scanner noise (39).

Stimuli were pseudorandomly presented in two blocks of 275 trials each. The order of block presentation was counterbalanced across subjects. Participants pressed a response key to indicate whether the two stimuli were congruent or incongruent, and participants did not respond during the rest trials. DMDX software (K. Forster, University of Arizona, Tucson) was used to present and control the timing of the stimuli (40).

Scan acquisition. Scanning was conducted on a 3-Tesla Bruker Medspec Avance S300 system by using a gradient-echo echoplanar imaging (EPI) sequence (TR = 3,000 ms, TE = 27.5 ms, TA = 1,100 ms, flip angle 85°, matrix size 64 × 64, FOV 20 × 20 cm, in plane resolution 3.125 mm × 3.125 mm, 21 oblique slices angled away from the eyes, 4-mm thick, with head coils, 101-kHz bandwidth, reconstruction based on a gradient-echo reference scan). Spoiled gradient recalled (SPGR) T1-weighted scans were acquired for anatomical localization. The data were preprocessed and analyzed by using SPM2 software (41) implemented in MATLAB (Mathworks, Natick, MA).

fMRI data analysis. Preprocessing included within-subject realignment, spatial normalization of the functional images to a standard EPI template, and spatial smoothing by using an anisotropic Gaussian kernel of 6 × 6 × 8 mm. Data for each subject were modeled with the general linear model by using the canonical hemodynamic response function with temporal derivatives. Parameter estimate images from each subject were combined into a group random effects analysis.

We rendered statistical parametric maps at $P = 0.001$ uncorrected and report cluster maxima with a random field corrected P value of $P < 0.05$ adjusted for the entire brain, unless otherwise stated. Effect sizes were analyzed with MARSBAR∥ by estimating the mean contrast values over all individual subjects. Unsmoothed data were used for the perirhinal cortex cluster because smoothing would have blurred activation from closely proximate neighboring regions into this structure. One-tailed, paired t tests compared the mean effect sizes. Montreal Neurological Institute coordinates are reported.

Behavioral Study. *Participants.* Twelve healthy mature control subjects (7 women; mean age = 62 years, SD = 10 years), two patients with HSE, and two LIFC patients participated in the study (see Fig. 2).

Materials. Shortened versions of the unimodal and crossmodal tasks from the efMRI study were administered. All conditions contained equal numbers of congruent and incongruent trials ($n = 30$ each) and, within each congruency condition, of living and nonliving stimuli ($n = 15$ in each congruency condition).

Procedure. The stimulus timing parameters were identical to the fMRI study except that participants were allowed 6 s to respond. The tasks were presented in a fixed, blocked order. Participants decided whether the two stimuli "went together" and responded by button press. DMDX software (40) controlled stimulus timing and presentation and collection of the accuracy and reaction time data.

Data analysis. Analyses were restricted to accuracy performances. MC's percentage correct performances were analyzed with repeated-measures ANOVAs on the means over all items (F_1), and the number correct performances were also summed over all participants for an analysis of items (F_2). Patient's accuracy performances were analyzed with scores and with χ^2 analyses.

Please see *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site, for additional information.

1. Desimone, R. & Ungerleider, L. G. (1989) in *Handbook of Neuropsychology*, eds. Boller, F. & Grafman, J. (Elsevier, Amsterdam), Vol. 2, pp. 267–299.
2. Mishkin, M., Ungerleider, L. G. & Macko, K. A. (1983) *Trends Neurosci.* **6,** 414–417.
3. Ungerleider, L. G. & Mishkin, M. (1982) in *Analysis of Visual Behavior*, eds. Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W. (MIT Press, Cambridge, MA), pp. 549–586.
4. Murray, E. A. & Bussey, T. J. (1999) *Trends Cogn. Sci.* **3,** 142–151.
5. Murray, E. A. & Richmond, B. J. (2001) *Curr. Opin. Neurobiol.* **11,** 188–193.
6. Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M. & Malach, R. (2001) *Cereb. Cortex* **11,** 287–297.
7. Moss, H. E., Rodd, J. M., Stamatakis, E. A., Bright, P. & Tyler, L. K. (2005) *Cereb. Cortex* **15,** 616–627.
8. Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M. & Moss, H. E. (2004) *J. Cogn. Neurosci.* **16,** 351–362.
9. Damasio, A. R. (1989) *Neural Comput.* **1,** 123–132.
10. Simmons, W. K. & Barsalou, L. W. (2003) *Cogn. Neuropsychol.* **20,** 451–486.
11. Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J. P. (2001) *Science* **292,** 290–293.
12. Rauschecker, J. P. & Tian, B. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 11800–11806.
13. Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J. & Rauschecker, J. P. (2001) *J. Cogn. Neurosci.* **13,** 1–7.
14. Semple, M. N. & Scott, B. H. (2003) *Curr. Opin. Neurobiol.* **13,** 167–173.
15. Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E. & Frackowiak, R. (1991) *Brain* **114,** 1803–1817.
16. Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R. & DeYoe, E. A. (2004) *Cereb. Cortex* **14,** 1008–1021.
17. Suzuki, W. A. & Amaral, D. G. (1994) *J. Comp. Neurol.* **350,** 497–533.
18. Murray, E. A., Malkova, L. & Goulet, S. (1998) in *Comparative Neuropsychology*, ed. Milner, A. D. (Oxford Univ. Press, Oxford), pp. 51–69.
19. Parker, A. & Gaffan, D. (1998) *Behav. Brain Res.* **93,** 99–105.
20. Desimone, R. & Gross, C. G. (1979) *Brain Res.* **178,** 363–380.
21. Beauchamp, M. S., Lee, K. E., Argall, B. D. & Martin, A. (2004) *Neuron* **41,** 809–823.
22. Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H. & Martin, A. (2004) *Nat. Neurosci.* **7,** 1190–1192.
23. Bruce, C., Desimone, R. & Gross, C. G. (1981) *J. Neurophysiol.* **46,** 369–384.
24. Ettlinger, G. & Wilson, W. A. (1990) *Behav. Brain Res.* **40,** 169–192.
25. McRae, K., de Sa, V. R. & Seidenberg, M. S. (1997) *J. Exp. Psychol. Gen.* **126,** 99–130.
26. Tyler, L. K. & Moss, H. E. (2001) *Trends Cogn. Sci.* **5,** 244–252.
27. Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. C. & Cohen, J. D. (1999) *Nature* **402,** 179–181.
28. Schall, J. D., Stuphorn, V. & Brown, J. W. (2002) *Neuron* **36,** 309–322.
29. Calvert, G. A. (2001) *Cereb. Cortex* **11,** 1110–1123.
30. Friston, K. J. (1997) *Hum. Brain Mapp.* **5,** 133–136.
31. Price, C. J., Warburton, E. A., Moore, C. J., Frackowiak, R. S. J. & Friston, K. J. (2001) *J. Cogn. Neurosci.* **13,** 419–429.
32. Higuchi, S. & Miyashita, Y. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 739–743.
33. Sakai, K. & Miyashita, Y. (1991) *Nature* **354,** 152–155.
34. Levy, D. A., Bayley, P. J. & Squire, L. R. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 6710–6715.
35. Suzuki, W. A. & Amaral, D. G. (1994) *J. Neurosci.* **14,** 1856–1877.
36. Lavenex, P. & Amaral, D. G. (2000) *Hippocampus* **10,** 420–430.
37. Parker, A. & Gaffan, D. (1998) in *Comparative Neuropsychology*, ed. Milner, A. D. (Oxford Univ. Press, Oxford), pp. 109–126.
38. Davies, R. R., Graham, K. S., Xuereb, J. H., Williams, G. B. & Hodges, J. R. (2004) *Eur. J. Neurosci.* **20,** 2441–2446.
39. Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M. & Bowtell, R. W. (1999) *Hum. Brain Mapp.* **7,** 213–223.
40. Forster, K. I. & Forster, J. C. (2003) *Behav. Res. Methods Instrum. Comput.* **35,** 116–124.
41. Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D. & Frackowiak, R. S. J. (1995) *Hum. Brain Mapp.* **2,** 189–210.