The language network is recruited but not required for non-verbal semantic processing

Anna A. Ivanova[1], Zachary Mineroff[1], Vitor Zimmerer[2], Nancy Kanwisher[1,3], Rosemary Varley[2], Evelina Fedorenko[1,3,4]

[1]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA*

[2]*Division of Psychology and Language Sciences, University College London, UK*

[3]*McGovern Institute for Brain Research, Massachusetts Institute of Technology, USA*

[4]*Department of Psychiatry, Massachusetts General Hospital, USA*

Abstract

Consistent with longstanding findings from neuropsychology, several brain regions in left frontal and temporal cortex respond robustly and selectively to language [1]. These regions, often referred to as the "language network", respond more strongly to meaningful stimuli (like words and sentences) than to stimuli devoid of meaning (like pseudowords and Jabberwocky sentences) [2]. But are these regions selectively recruited in processing *linguistic* meaning? Or do they instead store and/or process complex semantic information independent of its format (sentences or pictures)? In Experiment 1, we scanned participants with fMRI while they performed a semantic plausibility judgment task vs. a difficult perceptual control task on sentences and line drawings that describe/depict simple agent-patient interactions. We found that the language regions responded more strongly when participants performed the semantic task compared to the perceptual task, for both sentences and pictures (although sentences elicited overall stronger responses). Thus, healthy adults engage language regions when processing non-verbal meanings. But is this engagement necessary for understanding pictorial depictions of events? In Experiment 2, we tested two individuals with global aphasia, who have sustained massive damage to perisylvian language areas and display severe language difficulties, together with a group of control participants. Individuals with aphasia were at chance on a task of matching the sentences and pictures. However, they performed close to controls in assessing the plausibility of pictorial depictions of agent-patient interactions. Taken together, these results indicate that the left fronto-temporal language system is recruited but not necessary for processing complex non-verbal meanings.

Keywords: language, semantics, vision, fMRI, aphasia

Corresponding authors:
Anna Ivanova (annaiv@mit.edu); Evelina Fedorenko (evelina9@mit.edu)

Introduction

What is the relationship between thought and language? Do we use our language system for non-verbal thought? Can we think without language? Recent evidence from cognitive neuroscience suggests a surprisingly clear answer to these longstanding questions [3]. The fronto-temporal language system has been shown to respond very specifically in fMRI to linguistic stimuli, but not during arithmetic, music perception, working memory, and cognitive control tasks [2,4] and not during action or gesture perception [5,6]. Even stronger evidence comes from investigations of patients with massive left hemisphere damage and profound disruption of language capacity (global aphasia). These individuals can perform arithmetic and logical reasoning tasks, appreciate music, and think about another person's thoughts [7-10]. Thus, the available evidence suggests that language is neurally distinct from and not necessary for many aspects of thought. However, one domain where the division between linguistic and non-linguistic knowledge and processing remains elusive is semantics: the ability to store, represent, combine, and use previously acquired knowledge about the world.

Evidence from psycholinguistics indicates a tight integration of world knowledge about events with linguistic processing [11,12]. World knowledge helps resolve ambiguity during sentence processing, for instance, when inferring the meaning of *arrested* in "the criminal arrested..." vs. "the police arrested..." [13,14]. This inference process goes beyond the simple tracking of how often a given noun is the agent of a given verb: when inferring the agent of the verb "manipulate", people expected the noun phrase "the shrewd heartless gambler" more than the noun phrase "the young naïve gambler", an inference that depends on semantic knowledge rather than simple word co-occurrence [15]. Comprehenders also predict plausible patients based on the agent performing the action: the word "brakes" is easier to process following "the mechanic checked the..." vs. "the journalist checked the..." [16]. Taken together, these and related findings from the sentence processing literature suggest that people construct rich event representations while processing language, evaluate these representations against world knowledge, and use this knowledge incrementally as they interpret incoming words.

Further, much of our semantic knowledge about world entities and interactions between them plausibly comes from language. Although we can learn that cops tend to arrest criminals, but not vice versa, from visual information (live observation, movies, or comic books), evidence suggests that language can provide such information just as effectively. For example, congenitally blind individuals acquire word meanings that are remarkably similar to those of sighted adults [17,18] and that activate similar brain areas [19,20]. Furthermore, distributed semantic representations created through extensive training of computers on massive language corpora (e.g. [21]) can be used to successfully predict behavior across many semantic tasks [22,23] and even to decode the neural signature of visual stimuli [24,25]. Thus, language provides a rich source of world knowledge. However, does this mean that linguistic representations of meaning are inseparable from non-verbal representations? For instance, when we understand the meaning of, and extract world knowledge about, visual depictions of events, do we engage our language system? Could we do these things if we did not have access to linguistic resources?

In this paper, we synergistically combine fMRI and evidence from patients with brain damage to ask whether the language system is engaged during and/or necessary for the processing of non-verbal meaning. We focus on understanding of agent-patient relations ("who did what to whom") in pictorial depictions of events. Identification of thematic relations is critical to generating and understanding sentences [26,27], leading some neuroimaging studies to use it as a marker of linguistic processing [28]. At the same time, "agent" and "patient" are not exclusively linguistic notions; rather, they are ubiquitously used for understanding events, whether conveyed linguistically or non-linguistically [29,30], hinting at the possibility of abstract, cross-domain representation of thematic roles [31]. Neuroimaging can help resolve this debate by directly testing whether the language system supports the processing of visual stimuli that depict agent-patient interactions.

We used two different sources of evidence to test the role of the language system in non-verbal event processing: (1) functional neuroimaging of neurotypical participants, and (2) behavioral evidence from two individuals with global aphasia. In both experiments, participants evaluated the plausibility of events, presented either as sentences (neurotypicals only) or as line drawings. Plausibility was manipulated by varying the typicality of agent and patient roles (e.g., a cop arresting a criminal vs. a criminal arresting a cop). Performing this task requires going beyond mere semantic association (e.g. cop goes with criminal): in order to evaluate event plausibility, participants need to infer agent-patient roles from pictures and access stored knowledge about the typicality of those roles.

We find that language-responsive brain areas in neurotypical participants are active during the semantic plausibility task for both sentences and pictures, although the responses are overall lower for pictures. However, participants with global aphasia, in whom left-hemisphere language areas are severely damaged, still perform well on the picture plausibility task, suggesting that the language system is not required for understanding and accessing stored semantic information about visually presented events.

Experiment 1: Is the language system active during visual event comprehension?

In the first experiment, we presented twenty-one neurotypical participants with sentences and pictures describing agent-patient interactions that were either plausible or implausible (**Figure 1**), while the participants were undergoing an fMRI scan. Participants performed either a semantic judgment task on the sentences or pictures, or a difficulty-matched low-level perceptual control task on the identical stimuli, in a 2x2 blocked design (see SI for behavioral results). In separate blocks, participants were instructed to indicate either whether the stimulus was plausible or implausible (the semantic task) or whether the stimulus was moving slowly to the left or right (the perceptual task). The language system within each participant was identified using a separate functional language localizer task (sentences > nonwords contrast; [2]). We then measured the response of those regions to sentences and pictures during the semantic and perceptual tasks.

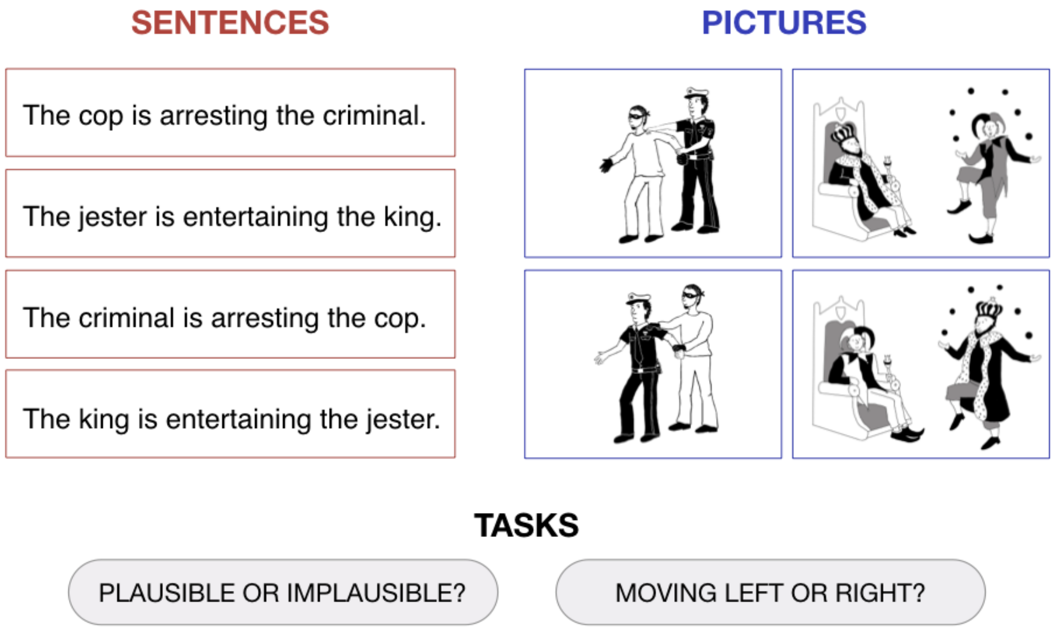

Figure 1. *Sample stimuli used for the experiment. For both pictures and sentences, participants were required to perform either a semantic plausibility judgment task or a control, perceptual task. The full set of materials is available on the paper website (https://osf.io/gsudr/).*

Although diverse non-linguistic tasks have been previously shown not to engage the language system [3], we here found that language-responsive regions were recruited more strongly during the semantic judgments on both sentences and pictures, compared to the perceptual control tasks on each (**Figure 2**). A linear mixed-effects model fitted to the data showed a significant effect of task (semantic > perceptual, $\beta = 1.07$, $p < .001$), a marginally significant effect of stimulus type (sentences > pictures, $\beta = .26$, $p = 0.06$), and no interaction between task and stimulus type ($\beta = .30$, *n.s.*). Because recent work has indicated that left angular gyrus (AngG) differs functionally from other language regions [32-36], we next repeated the analysis with the AngG excluded. This exploratory analysis confirmed the effect of task (semantic > perceptual, $\beta = .94$, $p < .001$) and additionally revealed an effect of stimulus type (sentences > pictures, $\beta = .38$, $p = 0.005$) and an interaction between task and stimulus type ($\beta = .41$, $p = 0.03$), showing that the core language regions respond more to sentences compared to pictures, but both stimulus types nonetheless elicit responses during the semantic task. We also verified that the semantic > perceptual effect for pictures in the language regions was caused by an increase in activity during the semantic task, rather than a decrease in activity during the perceptual task. Because we used the perceptual task on pictures as a baseline, we could estimate the level of activity within the language system during this condition by examining the regression intercept. The intercept was not significantly different from 0 ($\beta = .24$, *n.s.*), suggesting that the language system is not deactivated by the perceptual task. In sum, these results demonstrate that the language system is active during semantic processing of not only sentences, but also pictures.

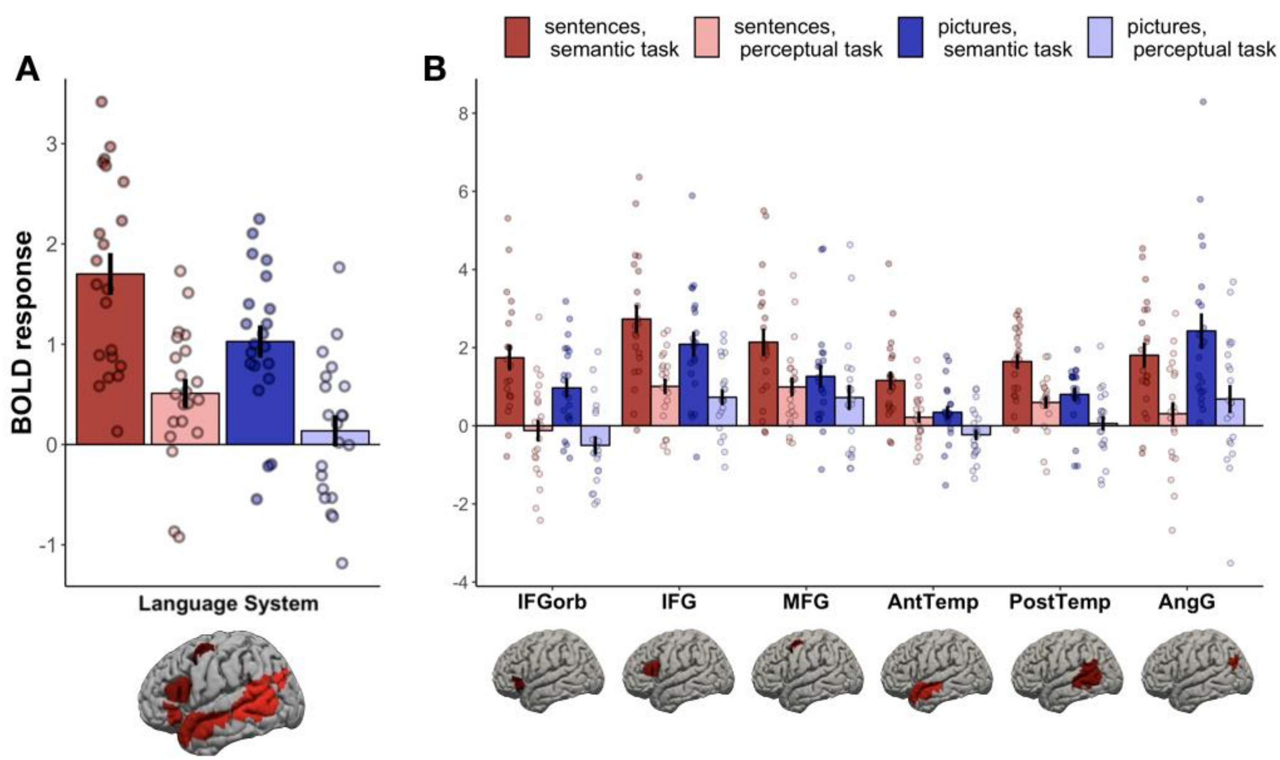

Figure 2. *BOLD response during the four task conditions within the language system (A) and each of the six language fROIs (B). All fROIs show higher activation for the semantic task than for the perceptual task, for both sentences and pictures (the task effect is significant for all fROIs except LMFG); anterior and posterior temporal fROIs also show a higher response for sentences than pictures. LIFGorb – left inferior frontal gyrus, pars orbitalis; LIFG – left inferior frontal gyrus, pars triangularis; LMFG – left middle frontal gyrus; LAntTemp – left anterior temporal cortex; LPostTemp – left posterior temporal cortex; LAngG – left angular gyrus. Within each parcel, the responses to the critical experiment conditions are extracted from the top 10% most language-responsive voxels (selected in each of the 24 individuals separately). Error bars indicate standard error of the mean.*

To investigate the different brain regions comprising the language network in more detail, we conducted follow-up analyses on individual fROIs' activity (Bonferroni-corrected for the number of regions). These revealed a significant difference in response to sentences compared to pictures in the left posterior temporal fROI (β = .54, p = .02) a marginally significant difference in the left anterior temporal fROI (β = .45, p = .07), and no significant differences in other fROIs. The semantic > perceptual effect was significant in five out of six fROIs (β = 0.58-1.74, p <= .007) and marginally significant in the middle frontal gyrus fROI (β = 0.54, p = 0.07). There was no interaction between task and stimulus type in any fROI. In sum, we show that all fROIs (with the possible exception of the middle frontal gyrus fROI) increase their activity during a semantic task when presented with both verbal and nonverbal stimuli.

We also performed a random effects whole-brain group analysis to determine the extent of activation outside the language fROIs (**Figure S1**). We found that the semantic > perceptual contrast for both sentences and pictures activates fronto-temporal regions that overlap with the language parcels; we have additionally found activation within the visual cortex, as well as within medial frontal and parietal cortices. Activations were observed bilaterally, with a stronger degree of lateralization observed for the sentence stimuli. The whole-brain analysis suggests that, while the left-lateralized language system is active during the semantic plausibility task, it may not be the only system that contributes to it.

The first experiment revealed that the language system is strongly and significantly recruited for semantic (compared to perceptual) processing of nonverbal stimuli - specifically, during accessing the meaning of line drawings depicting agent-patient interactions and relating them to stored world knowledge about these protagonists. This is the first study in which strong activation to a nonlinguistic task has been observed in independently-localized language regions. Next we ask: is the engagement of the language system necessary for extracting thematic roles from and world knowledge about visual scenes? To test the causal role of language regions in nonverbal semantics, we turn to behavioral evidence from global aphasia patients.

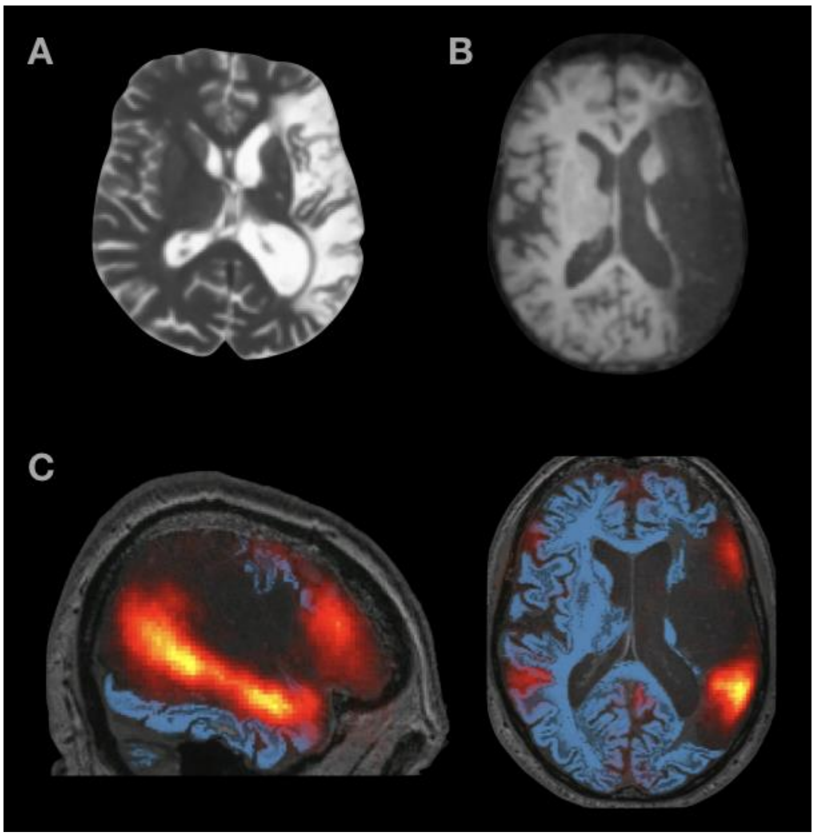

Figure 3. *Structural MRI images from (A) S.A. and (B) P.R. (C) Probabilistic language activation overlap map overlaid on top of P.R.'s structural MRI image. The heatmap values range from 0.01 (red) to 0.5 (yellow). As can be seen, the lesion covers most areas containing voxels that are likely to belong to the language network.*

Experiment 2: Is the language system necessary for visual event comprehension?

In the second experiment, we examined two individuals with global aphasia, a disorder characterized by severe impairments in both language production and language comprehension. Both patients had suffered large vascular lesions that resulted in extensive damage to left perisylvian cortex, including putative language fROIs (see **Figure 3** for MRI images, including an overlay of a probabilistic activation overlap map for the language system based on fMRI data and one of the participant's lesions) and consequent profound language impairment (see SI for language assessment scores). We measured patients' performance on two tasks: (1) the semantic task, identical to the picture version of the plausibility task from Experiment 1, but performed only on pictures, and (2) a sentence-picture matching task, during which participants saw a picture and a sentence where the agent and patient either matched the picture or were switched ("a cop is arresting a criminal" vs. "the criminal is arresting a cop"); participants then had to indicate whether or not the sentence matched the picture. For each task, patient performance was compared with the performance of 12 age-matched controls.

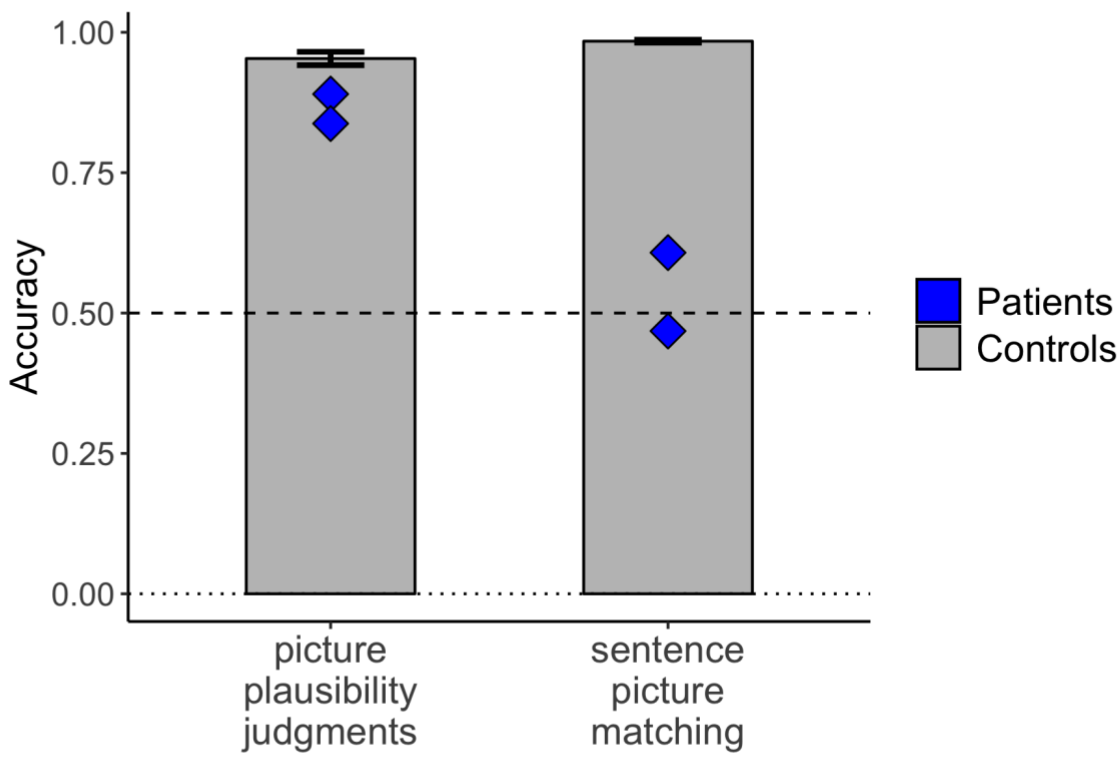

Figure 4. *Individuals with profound aphasia perform well on picture plausibility judgment task but fail on the sentence-picture matching task. Patient accuracies are indicated in blue; control data (N=12) are shown in gray bars. The dotted line indicates chance performance. Error bars indicate standard error of the mean.*

The results showed a clear difference in performance between the picture plausibility task and the sentence-picture matching task (**Figure 4**). Both aphasia patients and control participants performed well above chance when judging picture plausibility. Neurotypical controls had a mean accuracy of 95% (*SD* = 4.04%). Aphasia patients had mean accuracies of 88.7% (S.A.; 1.6 *SD* below average) and 83.8% (P.R.; 2.8 *SD* below average); exact binomial test showed that performance of both patients was above chance (S.A., $p < .001$, 95% CI [.79, .95]; P.R., $p < .001$, 95% CI [.73, .91]). Although their performance was slightly below the level of the controls, the data indicate that both patients were able to process complex semantic (agent-patient) relations in order to evaluate the plausibility of the depicted events.

In the sentence-picture matching task, control participants performed close to ceiling, with a mean accuracy of 98% (SD = 1.00%). In contrast, both patients were severely impaired: S.A. had a mean accuracy of 60.8% and P.R. had a mean accuracy of 46.8%. Exact binomial test showed that P.R.'s performance was at chance (*n.s.*, 95% CI [.39, .55]), while S.A.'s performance was above chance ($p = .009$, 95% CI [.52, .69]) but still much lower than that of the controls.

Taken together, the findings from Experiment 2 demonstrate the ability to understand thematic relations in pictures in the presence of profound linguistic impairment and inability to extract this information from the corresponding sentences. Specifically, these individuals are able to access stored world knowledge about depicted entities and flexibly combine this knowledge to evaluate real-life event plausibility.

Discussion

The role of the language system in nonverbal semantic processing has been long debated. Whereas previous studies have shown a functional dissociation between language and a wide range of other cognitive tasks [3], language-responsive regions have previously been implicated in high-level semantic processing [37,38]. Here, we demonstrate that left-hemisphere language regions are, in fact, active during semantic processing of pictures depicting agent-patient interactions. Importantly, however, this activation does not appear to be essential for semantic judgments, given that the two tested individuals with global aphasia, who lack most of their left-hemisphere language network and have severe language impairments, can still evaluate the plausibility of visually presented events.

Determining whether a picture depicting an interaction between two actors is plausible or not is a multi-component process, requiring the person to identify (1) relevant actors, (2) the action taking place between them, (3) the role that each actor is performing (agent vs. patient), and (4) the likelihood that a given actor would be the agent/patient of the relevant action. Whereas the first three components can, at least in part, be attributed to modality-specific perceptual processing, establishing plausibility cannot be solely attributed to vision. In order to decide whether a cop arresting a criminal is more likely than a criminal arresting a cop, participants would need to evaluate the plausibility of these two interactions by drawing on their world knowledge. We demonstrate that this highly abstract process can proceed even when the language system is severely impaired, thus providing further evidence that semantics and language are neurally distinct.

Linguistic and non-linguistic meanings are distinct. The dissociation observed between language-based and vision-based semantic processing indicates that these functions are supported by different brain networks. This result accords with the fact that both non-human animals and preverbal infants are capable of complex event processing [39,40] and that specialized neural mechanisms, distinct from the language system, have been associated with visual understanding of actions [41-45] and interactions between animate and/or inanimate entities [46-48].

Yet because many neuroimaging studies that probe semantic processing use sentences or words as stimuli (cf. [49,50]), the distinction between linguistic and non-linguistic meaning is often lost. As a result, most research on semantic processing in the brain points to regions that are also implicated in language processing, such as anterior temporal regions (e.g., [51]), left inferior frontal cortex (e.g., [52]), and left angular gyrus (e.g., [53,54]). However, studies that aim to decode meanings of words [55] or sentences [56,57] from neuroimaging data show that linguistic meaning can be found in a widely distributed set of voxels rather than a focal hub. [57] specifically point out that many of the voxels that lend themselves to sentence decoding can be found outside the language regions. Moreover, our whole-brain analysis (**Figure S1**) shows that a semantic task elicits activations bilaterally, and so do neuroimaging studies that use semantically complex stimuli, such as narratives [58-59]. These results, as well as the fact that patients with extensive left-hemisphere damage can understand complex pictorial stimuli, further emphasize that fronto-temporal left hemisphere regions are not a critical contributor to semantic processing.

Future research should test the extent to which semantic processing of interactions between two animate actors differs from other forms of semantic processing. Many of our stimuli depict social interactions, an event type that occupies a privileged position in perception. Processing of human interactions is automatic and facilitates recognition of participating actors [30,60-63]. Some have argued that identifying an agent in a scene automatically initiates event structure building, a process that involves accessing world knowledge in order to predict possible upcoming actions [29]. Furthermore, visual presentation of social interaction scenes activates multiple processing systems in the human brain [64], including interaction-specific brain regions in both humans [47] and monkeys [65], and viewing intentional actions activates areas in right posterior superior temporal sulcus [66,67]. These findings point to the possibility that our patients might have been recruiting regions dedicated to processing social interactions to perform the plausibility task on the pictures.

Visual processing of agent-patient interactions does not rely on the language system. Previous experiments attempting to localize brain region(s) involved in thematic role processing have mostly pointed to the left temporal lobe. In a neuroimaging study that investigated the neural basis of agent/patient identification using sentences, Frankland and Greene [68] isolated distinct areas in left superior temporal sulcus (STS) that are sensitive to the identity of the agent vs. patient. Another study [69] found that the same (or a nearby) STS regions also contained information about event participant roles in videos depicting agent-patient interactions. However, the latter study also identified a number of other regions that were sensitive to thematic roles, including clusters in the right posterior middle temporal gyrus and right angular gyrus. Our results suggest that thematic role representations extracted from sentences and visual scenes are, in fact, distinct: even though visual presentation of events activates the language regions in healthy individuals, their involvement is not crucial for correctly identifying agents and patients in a given scene.

An earlier neuropsychological study [70] investigated patients with various left-hemisphere lesions and found that lesions to mid-STS led to difficulties in extracting thematic role information from both sentences and pictures, a finding that is inconsistent with our results. However, deficits in processing agent-patient relationships in pictures in Wu et al.'s study were additionally associated with lesions in anterior superior temporal gyrus, supramarginal gyrus, and inferior frontal cortex. Given that, in our work, the two patients with extensive left-hemisphere damage were able to identify thematic roles from pictures, future research should closely investigate the effects of damage to *specific areas* within the left hemisphere on nonverbal semantic processing.

The language system is active during a visual task – even though its engagement is not essential. In Experiment 1, we observed activation within the language system in response to an event semantics task performed on pictures. This activation, however, does not seem to be essential for visual event processing, since patients with profound aphasia were mostly able to perform the task. What does the observed activity reflect then? One possibility is that subjects covertly recode pictorial stimuli into verbal form. Trueswell and Papafragou [71] demonstrated covert language encoding during a picture memorization task, while Greene and Fei-Fei [72] showed a delay in categorizing a word (as object vs. scene) if that word was overlaid on top of an incongruent image. On the other hand, psycholinguistic work that uses verbal descriptions of an item set ("visual world" studies) indicates that participants do not automatically generate verbal labels for visually presented objects [73,74; cf. 75]. Furthermore, Potter and colleagues [76, 77] have shown that verbal recoding of pictorial information is relatively slow and can only occur after the concept of the picture is retrieved, suggesting that online construction of a verbal picture description is effortful. Our stimuli depicted complex multi-actor events, making verbal recoding even more effortful and, therefore, less likely to occur during a fast-paced task. Overall, however, the verbal recoding hypothesis cannot be excluded and requires further investigation.

Another possibility is that the language system can meaningfully contribute to the task, even if its engagement is not required. Recent successes in deriving semantic representations using word embedding models have placed a strong emphasis on linguistic information as a source of world knowledge, leading to the development of language-based models that can perform "conceptual" tasks, such as inference, paraphrasing or question answering [78,79, among others]. Even simple n-gram models can determine the probability of certain events by, e.g., estimating the probability that the phrase “is arresting” directly follows “cop” vs. “criminal”. Given that semantic information derived from language is distinct from perception-based world knowledge [80] and that people can flexibly use both depending on task demands [81], it is possible that the language system in neurotypicals provides an additional source of information when determining event plausibility.

Regardless of the explanation, the non-causal nature of the language system activation during a nonverbal task has important implications for semantics research. A significant body of work has aimed to isolate “amodal” representations of concepts by investigating the overlap between regions active during verbal and nonverbal presentations of a stimulus [82-89]. Most of these overlap-based studies have attributed semantic processing to frontal, temporal, and/or parietal regions within the left hemisphere – which is unsurprising given that the language system in most people is located within these areas (and so the overlap could not have been found elsewhere). Our work, however, demonstrates that left-lateralized activity evoked by both language and visual stimuli is not essential for nonverbal semantic processing, and that conceptual information about events (specifically, plausibility of actors performing certain actions) persists even when most fronto-temporo-parietal language regions are damaged.

Overall, our study emphasizes the inherent correlational nature of functional neuroimaging findings and the importance of causal evaluation of the theories of cognitive and neural architecture. We find that although the language system is active during a semantic plausibility judgment task on agent-patient interaction in pictures, individuals with a severely disrupted language system still perform well above chance on that task. Our results indicate that agent-patient interactions and access to world knowledge do not depend on the language system and highlight the necessity to further study semantic processes that operate on non-linguistic input.

Materials and Methods

Experiment 1

Participants. Twenty-four participants took part in the fMRI experiment (11 female, mean age = 25 years, SD = 5.2). The participants were recruited from MIT and the surrounding Cambridge/Boston, MA, community and paid for their participation. All were native speakers of English, had normal hearing and vision, and no history of language impairment. All were right-handed (as assessed by Oldfield’s handedness questionnaire [90], or self-report). Two participants had low behavioral accuracy scores (<60%), and one had right-lateralized language regions (as evaluated by the language localizer task; see below); they were excluded from the analyses, which were therefore based on data from 21 participants. The protocol for the study was approved by MIT’s Committee on the Use of Humans as Experimental Subjects (COUHES). All participants gave written informed consent in accordance with protocol requirements.

Design, materials, and procedure. All participants completed a language localizer task aimed at identifying language-responsive brain regions [2] and the critical picture/sentence plausibility task.

The localizer task was conducted in order to identify brain regions within individual participants that selectively respond to language stimuli. During the task, participants read sentences (e.g., NOBODY COULD HAVE PREDICTED THE EARTHQUAKE IN THIS PART OF THE COUNTRY) and lists of unconnected, pronounceable nonwords (e.g., U BIZBY ACWORRILY MIDARAL MAPE LAS POME U TRINT WEPS WIBRON PUZ) in a blocked design. Each stimulus consisted of twelve words/nonwords. For details of how the language materials were constructed, see [2]. The materials are available at http://web.mit.edu/evelina9/www/funcloc/funcloc_localizers.html. The sentences > nonword-lists contrast has been previously shown to reliably activate left-lateralized fronto-temporal language processing regions and to be robust to changes in the materials, task, and modality of presentation [2,91-92]. Stimuli were presented in the center of the screen, one word/nonword at a time, at the rate of 450 ms per word/nonword. Each stimulus was preceded by a 100 ms blank screen and followed by a 400 ms screen showing a picture of a finger pressing a button, and a blank screen for another 100 ms, for a total trial duration of 6 s. Participants were asked to press a button whenever they saw the picture of a finger pressing a button. This task was included to help participants stay alert and awake. Condition order was counterbalanced across runs. Experimental blocks lasted 18 s (with 3 trials per block), and fixation blocks lasted 14 s. Each run (consisting of 5 fixation blocks and 16 experimental blocks) lasted 358 s. Each participant completed 2 runs.

The picture plausibility task included two types of stimuli: (1) black-and-white line drawings depicting plausible and implausible agent-patient interactions (created by an artist for this study), and (2) simple sentences describing the same interactions. Sample stimuli are shown in Figure 1, and a full list of materials is available on the paper website (https://osf.io/gsudr/). Forty plausible-implausible pairs of pictures, and forty plausible-implausible pairs of corresponding sentences were used. The full set of materials was divided into two lists, such that List 1 used plausible pictures and implausible sentences for odd-numbered items, and implausible pictures and plausible sentences for even-numbered items, and List 2 did the opposite. Thus, each list contained either a picture or a sentence version of any given event. Stimuli were presented in a blocked design (each block included either pictures or sentences) and were moving either to the right or to the left for the duration of stimulus presentation. At the beginning of each block, participants were told which task they would have to perform next: the semantic or the perceptual one. The semantic task required them to indicate whether the depicted/described event is plausible or implausible, by pressing one of two buttons. The perceptual task required them to indicate the direction of stimulus movement (right or left). To ensure that participants always perform the right task, a reminder about the task and the response buttons ("plausible=1/implausible=2", or "moving right=1/left=2") was visible in the lower right-hand corner of the screen for the duration of the block. Each stimulus (a picture or a sentence) was presented for 1.5 s, with 0.5 s intervals between stimuli. Each block began with a 2-second instruction screen to indicate the task, and consisted of 10 trials, for a total duration of 22 s. Trials were presented with a constraint that the same response (plausible/implausible in the semantic condition, or right/left in the perceptual condition) did not occur more than 3 times in a row. Each run consisted of 3 fixation blocks and 8 experimental blocks (2 per condition: semantic task – pictures, semantic task – sentences, perceptual task – pictures, perceptual task - sentences) and lasted 242 s (4 min 2 s). The order of conditions was palindromic and varied across runs and participants. Each participant completed 2 runs.

fMRI data acquisition. Structural and functional data were collected on the whole-body, 3 Tesla, Siemens Trio scanner with a 32-channel head coil, at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted structural images were collected in 176 sagittal slices with 1mm isotropic voxels (TR=2,530ms, TE=3.48ms). Functional, blood oxygenation level dependent (BOLD), data were acquired using an EPI sequence (with a 90º flip angle and using GRAPPA with an acceleration factor of 2), with the following acquisition parameters: thirty-one 4mm thick near-axial slices acquired in the interleaved order (with 10% distance factor), 2.1mm×2.1mm in-plane resolution, FoV in the phase encoding (A>>P) direction 200mm and matrix size 96mm×96mm, TR=2000ms and TE=30ms. The first 10s of each run were excluded to allow for steady state magnetization.

fMRI data preprocessing. MRI data were analyzed using SPM5 (see SI for a note on software usage) and custom MATLAB scripts (available in the form of an SPM toolbox from http://www.nitrc.org/projects/spm_ss). Each participant's data were motion corrected and then normalized into a common brain space (the Montreal Neurological Institute (MNI) template) and resampled into 2mm isotropic voxels. The data were then smoothed with a 4mm FWHM Gaussian filter and high-pass filtered (at 200s). Effects were estimated using a General Linear Model (GLM) in which each experimental condition was modeled with a boxcar function (modeling entire blocks) convolved with the canonical hemodynamic response function (HRF).

Defining functional regions of interest (fROIs). The critical analyses were restricted to individually defined language fROIs (functional regions of interest). These fROIs were defined using the Group-constrained Subject-Specific (GSS) approach (Fedorenko et al. 2010; Julian et al. 2012) where a set of spatial parcels is combined with each individual subject's localizer activation map, to constrain the definition of individual fROIs. The parcels mark the expected gross locations of activations for a given contrast based on prior work and are sufficiently large to encompass the extent of variability in the locations of individual activations. Here, we used a set of six parcels derived from a group-level probabilistic activation overlap map for the sentences > nonwords contrast in 220 participants. These parcels included two regions in the left inferior frontal gyrus (LIFG, LIFGorb), one in the left middle frontal gyrus (LMFG), two in the left temporal lobe (LAntTemp and LPostTemp), and one extending into the angular gyrus (LAngG). The parcels are available for download from https://evlab.mit.edu/funcloc. Within each parcel, we selected the top 10% most responsive voxels, based on the *t* values for the sentences > nonwords contrast (see Figure 1 in [33] or Figure 1 in [91], for sample fROIs). Individual-level fROIs defined in this way were then used for subsequent analyses that examined the behavior of the language network during the critical picture/sentence plausibility task.

Examining the functional response profiles of the language fROIs. For each language fROI in each participant, we averaged the responses across voxels to get a value for each of the four critical task conditions (semantic task – pictures, semantic task – sentences, perceptual task – pictures, perceptual task – sentences). We then ran a linear mixed-effect regression model with two fixed effects (stimulus type and task) and two random effects (participant and fROI). Model formula can be found in the SI. Planned follow-up comparisons examined response to sentences and pictures during the semantic task within each fROI; the results were Bonferroni-corrected. Model formulae can be found in the SI. The analysis was run using the *lmer* function from the *lme4* R package [93]; statistical significance of the effects was evaluated using the *lmerTest* package [94].

Experiment 2

Participants. Two patients with global aphasia took part in the study. Both had large lesions that had damaged the left inferior frontal gyrus, the inferior parietal lobe (supramarginal and angular gyri) and the superior temporal lobe. At the time of testing, they were 68 and 70 years old respectively. S.A. was pre-morbidly right-handed; P.R. was pre-morbidly left-handed, but a left hemisphere lesion that resulted in profound aphasia indicated that he, like most left-handers, was left-hemisphere dominant for language [95]. Neither participant presented with visual impairments.

The patients were classified as severely agrammatic and had severe difficulties with language tasks, ranging from single word production to sentence comprehension. Nonetheless, their performance on most non-verbal reasoning tasks, such as Pyramids and Palm Trees Test (PPT, [96]), was not impaired. Detailed information on patient assessment can be found in the SI.

We also tested two sets of neurotypical control participants, one for the semantic task, and one for the language task. The semantic task control participants were 12 healthy participants (7 females) ranging in age from 58 to 78 years (mean age 65.5 years). The language task control participants were 12 healthy participants (5 females) ranging in age from 58 to 78 years (mean age 64.7 years). All healthy participants had no history of speech or language disorders, neurological diseases or reading impairments. They were all native English speakers, and had normal, or corrected-to-normal, vision.

Participants undertook the experiments individually, in a quiet room. An experimenter was present throughout the testing session. The stimuli were presented on an Acer Extensa 5630G laptop, with the experiment built using DMDX [97].

Semantic Task: Picture plausibility judgments. The same picture stimuli were used as those in Experiment 1 (see Figure 1), plus one additional plausible-implausible pair of pictures (which was omitted from the fMRI experiment to have a total number of stimuli be divisible by four, for the purposes of grouping materials into blocks and runs), for a total of 82 pictures (41 plausible-implausible pairs). Four of the 82 pictures were used as training items (see below).

The stimuli were divided into 2 sets, with an equal number of plausible and implausible pictures; each plausible-implausible pair was split across the 2 sets, to minimize repetition of the same event participants within a set. The order of the trials was randomized within each set, so that each participant saw the pictures in a different sequence. A self-timed break was placed between the two sets.

Prior to the experiment, participants were shown two pairs of pictures, which acted as training items. The pairs consisted of one plausible and one implausible event. They were given clear instructions to focus on the relationship between the two characters and assess whether they thought the interaction was plausible, in adherence with normal expectations, or implausible, somewhat peculiar and at odds with expectations. They were asked to press a green tick (the left button on the mouse) if they thought the picture depicted a plausible event, and a red cross (the right button on the mouse) if they thought the picture depicted an implausible event. They were asked to do so as quickly and accurately as possible. The pictures appeared for a maximum of 8 seconds, with the inter-stimulus interval of 2 seconds. Accuracies and reaction times were recorded. Participants had the opportunity to ask any questions, and the instructions for participants with aphasia were supplemented by gestures to aid comprehension of the task. Participants had to indicate that they understood the task prior to starting.

Language Task: Sentence-picture matching. The same 82 pictures were used as in the plausibility judgment experiment. In this control task, a sentence was presented below each picture that either described the picture correctly (e.g., “the policeman is arresting a criminal” for the first sample picture in Figure 1) or had the agent and patient switched (“the criminal is arresting the policeman"[1]). Simple active subject-verb-object sentences were used. Combining each picture with a matching and a mismatching sentence resulted in 164 trials in total.

For the control participants, the trials were split into two sets of 82, with an equal number of plausible and implausible pictures, as well as an equal number of matches and mismatches in each set. In order to avoid tiring the participants with aphasia, the experiment was administered across two testing sessions each consisting of two sets of 41 stimuli and occurring within the same week. For both groups, the order of the trials was randomized separately for each participant, and no pictures belonging to the same "set" (i.e., an event involving a cop and a criminal) appeared in a row. A self-timed break was placed between the two sets.

Prior to the experiment, participants were told that they would see a series of pictures with accompanying sentences, and their task was to decide whether the sentence matched the event depicted. They were asked to press a green tick (the left button on the mouse) if they thought the sentence matched the picture, and a red cross (the right button on the mouse) if they thought the sentence did not match the picture. They were asked to do so as quickly and accurately as possible. The picture + sentence combinations appeared for a maximum of 25 seconds, with the inter-stimulus interval of 2 seconds. Accuracies and reaction times were recorded. As in the critical task, participants had the opportunity to ask any questions, and the instructions for participants with aphasia were supplemented by gestures.

Data analysis. We used exact binomial test to test whether patients' performance on either task was significantly above chance. We excluded all items with reaction times and/or accuracies outside 3 standard deviations of the control group mean (4 items for the semantic task and 11 items for the sentence-picture matching task).

Estimating the damage to the language network in patients with aphasia. In order to estimate the extent of the damage to the language network, we combined available structural MRI of one patient with aphasia with a probabilistic activation overlap map of the language network. The map was created by overlaying thresholded individual activation maps for the language localizer contrast (sentences > nonwords, as described in Experiment 1) in 220 healthy participants. The maps were thresholded at the $p < 0.001$ whole-brain uncorrected level, binarized, and overlaid in the common space, so that each voxel contains information on the proportion of participants showing a significant language localizer effect (see [98] for more details). The map can be downloaded from the Fedorenko lab website (https://evlab.mit.edu/funcloc/download-overlap_maps). An overlay of this probabilistic map onto P.R.'s structural scan is shown in Figure 3.

Acknowledgments

We would like to acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT, and its support team (Steve Shannon, Atsushi Takahashi, and Sheeba Arnold). We thank Birgit Zimmerer for creating the picture stimuli used in both experiments, Chloe Bustin for stimuli norming, Lily Jordan for help with the behavioral piloting of the fMRI experiment, EvLab members for their help with fMRI data collection. E.F. was supported by NIH awards R00-HD057522, R01-DC016607, and R01-DC016950, and by a grant from the Simons Foundation to the

[1] For this stimulus sentence, "cop" was used in Experiment 1 (conducted in the US) and "policeman" was used in Experiment 2 (conducted in the UK).

Simons Center for the Social Brain at MIT. R.V. was supported by Arts and Humanities Research Council and Alzheimer's Society awards.

References

1. Fedorenko E, Behr MK, Kanwisher N (2011) Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 108(39):16428-33.

2. Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N (2010) New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J. Neurophysiol.* 104(2):1177-94.

3. Fedorenko E, Varley R (2016) Language and thought are not the same thing: evidence from neuroimaging and neurological patients. *Ann. N. Y. Acad. Sci.* 1369(1):132-153.

4. Monti MM, Parsons LM, Osherson DN (2012) Thought beyond language: Neural dissociation of algebra and natural language. *Psychol. Sci.* 23(8):914-22.

5. Pritchett BL, Hoeflin C, Koldewyn K, Dechter E, Fedorenko E (2018) High-level language processing regions are not engaged in action observation or imitation. *J. Neurophysiol.* 120(5):2555-70.

6. Jouravlev O, Zheng D, Balewski Z, Pongos A, Levan Z, Goldin-Meadow S, Fedorenko E (in press) Speech-accompanying gestures are not processed by the language-processing mechanisms, *Neuropsychologia*, doi: https://doi.org/10.1016/j.neuropsychologia.2019.107132.

7. Basso A, Capitani E (1985) Spared musical abilities in a conductor with global aphasia and ideomotor apraxia. *J. Neurol. Neurosurg. Psychiatry* 48(5):407-12.

8. Luria AR, Tsvetkova LS, Futer DS (1965) Aphasia in a composer. *J. Neurol. Sci.* 2(3):288-92.

9. Varley RA, Klessinger NJ, Romanowski CA, Siegal M (2005) Agrammatic but numerate. *Proc. Natl. Acad. Sci. U.S.A.* 102(9):3519-24.

10. Varley R, Siegal M (2000) Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Curr. Biol.* 10(12):723-6.

11. Hagoort P, van Berkum J (2007) Beyond the sentence given. *Philos. Trans. Royal Soc. B.* 362(1481):801-11.

12. Matsuki K, Chow T, Hare M, Elman JL, Scheepers C, McRae K (2011) Event-based plausibility immediately influences on-line language comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 37(4):913.

13. McRae K, Spivey-Knowlton MJ, Tanenhaus MK (1998) Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *J. Mem. Lang.* 38(3):283-312.

14. Traxler MJ, Morris RK, Seely RE (2002) Processing subject and object relative clauses: Evidence from eye movements. *J. Mem. Lang.* 47(1):69-90.

15. McRae K, Ferretti TR, Amyote L (1997) Thematic roles as verb-specific concepts. *Lang. Cogn. Process.* 12(2-3):137-76.

16. Bicknell K, Elman JL, Hare M, McRae K, Kutas M (2010) Effects of event knowledge in processing verbal arguments. *J. Mem. Lang.* 63(4):489-505.

17. Landau B, Gleitman LR (1985) *Cognitive science series, 8. Language and experience: Evidence from the blind child.* Harvard University Press.

18. Marmor GS (1978). Age at onset of blindness and the development of the semantics of color names. *J. Exp. Child. Psychol.* 25(2):267–278.

19. Bedny M, Caramazza A, Grossman E, Pascual-Leone A, Saxe R (2008) Concepts are more than percepts: the case of action verbs. *J. Neurosci*. 28(44):11347-11353.

20. Bedny M, Caramazza A, Pascual-Leone A, Saxe R (2011). Typical neural representations of action verbs develop without vision. *Cereb. Cortex* 22(2):286-293.

21. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. *Adv. Neural. Inf. Process. Syst.* 26:3111-3119.

22. Grand G, Blank IA, Pereira F, Fedorenko E (2018) Semantic projection: recovering human knowledge of multiple, distinct object features from word embeddings. *arXiv preprint* arXiv:1802.01241.

23. Pereira F, Gershman S, Ritter S, Botvinick M (2016) A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cogn. Neuropsychol.* 33(3-4):175-90.

24. Güçlü U, van Gerven MA (2015) Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* 35(27):10005-14.

25. Nishida S, Nishimoto S (2018) Decoding naturalistic experiences from human brain activity via distributed representations of words. *NeuroImage* 180:232-42.

26. Fillmore CJ (2003) *Form and meaning in language.* CSLI Publications, 121.

27. Jackendoff RS (1974) *Semantic Interpretation in Generative Grammar.* MIT Press.

28. Monti MM, Parsons LM, Osherson DN (2009) The boundaries of language and thought in deductive inference. *Proc. Natl. Acad. Sci. U.S.A.* 106(30):12554-9.

29. Cohn N, Paczynski M (2013) Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cogn. Psychol.* 67(3):73-97.

30. Hafri A, Trueswell JC, Strickland B (2018) Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition* 175:36-52.

31. Spelke ES, Kinzler KD (2007) Core knowledge. *Dev. Sci.* 10(1):89-96.

32. Amit E, Hoeflin C, Hamzah N, Fedorenko E (2017) An asymmetrical relationship between verbal and visual thinking: Converging evidence from behavior and fMRI. *NeuroImage* 152:619-627.

33. Blank I, Kanwisher N, Fedorenko E (2014) A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J. Neurophysiol.* 112(5):1105-18.

34. Mineroff Z, Blank IA, Mahowald K, Fedorenko E (2018) A robust dissociation among the language, multiple demand, and default mode networks: evidence from inter-region correlations in effect size. *Neuropsychologia* 119:501-11.

35. Paunov AM, Blank IA, Fedorenko E (2019) Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *J. Neurophysiol.* 121(4):1244-65.

36. Pritchett BL, Hoeflin C, Koldewyn K, Dechter E, Fedorenko E (2018) High-level language processing regions are not engaged in action observation or imitation. *J. Neurophysiol.* 120(5):2555-70.

37. Anderson AJ, Lalor EC, Lin F, Binder JR, Fernandino L, Humphries CJ, Conant LL, Raizada RD, Grimm S, Wang X (2018) Multiple regions of a cortical network commonly encode the meaning of words in multiple grammatical positions of read sentences. *Cereb. Cortex.* 29(6):2396-411.

38. Binder JR, Desai RH, Graves WW, Conant LL (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19(12):2767-2796.

39. Spelke E (1976) Infants' intermodal perception of events. *Cogn. Psychol.* 8(4):553-560.

40. Seed A, Tomasello M (2010) Primate cognition. *Top. Cogn. Sci.* 2(3):407-19.

41. Fang Y, Chen Q, Lingnau A, Han Z, Bi Y (2016) Areas recruited during action understanding are not modulated by auditory or sign language experience. *Front. Human Neurosci.* 10:94.

42. Häberling IS, Corballis PM, Corballis MC (2016) Language, gesture, and handedness: Evidence for independent lateralized networks. *Cortex* 82:72-85.

43. Kable JW, Kan IP, Wilson A, Thompson-Schill SL, Chatterjee A (2005) Conceptual representations of action in the lateral temporal cortex. *J. Cogn. Neurosci.* 17(12):1855-70.

44. Tarhan L, Konkle T (2019) Sociality and Interaction Envelope Organize Visual Action Representations. *bioRxiv* 1:618272.

45. Tranel D, Kemmerer D, Adolphs R, Damasio H, Damasio AR (2003) Neural correlates of conceptual knowledge for actions. *Cogn. Neuropsychol.* 20(3-6):409-32.

46. Fischer J, Mikhael JG, Tenenbaum JB, Kanwisher N (2016) Functional neuroanatomy of intuitive physical inference. *Proc. Natl. Acad. Sci. U.S.A.* 113(34):E5072-81.

47. Isik L, Koldewyn K, Beeler D, Kanwisher N (2017) Perceiving social interactions in the posterior superior temporal sulcus. *Proc. Natl. Acad. Sci. U.S.A.* 114(43):E9145-52.

48. Walbrin J, Downing P, Koldewyn K (2018) Neural responses to visually observed social interactions. 112:31-9.

49. Chertkow H, Bub D, Deaudon C, Whitehead V (1997) On the status of object concepts in aphasia. *Brain Lang.* 58(2):203-32.

50. Peelen MV, Caramazza A (2012) Conceptual object representations in human anterior temporal cortex. *J. Neurosci.* 32(45):15728-36.

51. Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Rev. Neurosci.* 8(12):976.

52. Hagoort P, van Berkum J (2007) Beyond the sentence given. *Philos. Trans. Royal Soc. B.* 362(1481):801-11.

53. Binder JR, Desai RH (2011) The neurobiology of semantic memory. *Trends Cogn. Sci.* 15(11):527-536.

54. Desai RH, Binder JR, Conant LL, Mano QR, Seidenberg MS (2011) The neural career of sensory-motor metaphors. *J. Cogn. Neurosci.* 23(9):2376-86.

55. Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. *Science.* 320(5880):1191-5.

56. Anderson AJ, Binder JR, Fernandino L, Humphries CJ, Conant LL, Aguilar M, Wang X, Doko D, Raizada RD (2016) Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. *Cereb. Cortex.* 27(9):4379-95.

57. Pereira F, Lou B, Pritchett B, Ritter S, Gershman SJ, Kanwisher N, Botvinick M, Fedorenko E (2018) Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.* 9(1):963.

58. Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 31(8):2906-15.

59. Huth AG, De Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532(7600):453.

60. Hafri A, Papafragou A, Trueswell JC (2013) Getting the gist of events: Recognition of two-participant actions from brief displays. *J. Exp. Psychol. Gen.* 142(3):880.

61. Neri P, Luu JY, Levi DM (2006) Meaningful interactions can enhance visual discrimination of human agents. *Nat. Neurosci.* 9(9):1186.

62. Papeo L, Stein T, Soto-Faraco S (2017) The two-body inversion effect. *Psychol. Sci.* 28(3):369-79.

63. Webb A, Knott A, MacAskill MR (2010) Eye movements during transitive action observation have sequential structure. *Acta Psychol.* 133(1):51-6.

64. Quadflieg S, Koldewyn K (2017) The neuroscience of people watching: how the human brain makes sense of other people's encounters. *Ann. N. Y. Acad. Sci.* 1396(1):166-82.

65. Sliwa J, Freiwald WA (2017) A dedicated network for social interaction processing in the primate brain. *Science* 356(6339):745-9.

66. Saxe R, Xiao DK, Kovacs G, Perrett DI, Kanwisher N (2004) A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42(11):1435-46.

67. Shultz S, Lee SM, Pelphrey K, McCarthy G (2010) The posterior superior temporal sulcus is sensitive to the outcome of human and non-human goal-directed actions. *Soc. Cogn. Affect. Neurosci.* 6(5):602-11.

68. Frankland SM, Greene JD (2015) An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences, 112*(37), 11732-11737.

69. Wang J, Cherkassky VL, Yang Y, Chang KMK, Vargas R, Diana N, Just MA (2016). Identifying thematic roles from neural representations measured by functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. U.S.A*. 33(3-4):257-264.

70. Wu DH, Waller S, Chatterjee A (2007) The functional neuroanatomy of thematic role and locative relational knowledge. *J. Cogn. Neurosci.*, 19(9):1542-1555.

71. Trueswell JC, Papafragou A (2010) Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *J. Mem. Lang.* 63(1):64-82.

72. Greene MR, Fei-Fei L (2014) Visual categorization is automatic and obligatory: Evidence from Stroop-like paradigm. *J. Vis.* 14(1):14-14.

73. Dahan D, Magnuson JS, Tanenhaus MK (2001) Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cogn. Psychol.* 42(4):317-367.

74. Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D (2003) The time course of spoken word learning and recognition: studies with artificial lexicons. *J. Exp. Psychol. Gen.* 132(2):202.

75. Meyer AS, Belke E, Telling AL, Humphreys GW (2007). Early activation of object names in visual search. *Psychon. Bull. Rev*. 14(4):710-716.

76. Potter MC, Faulconer BA (1975). Time to understand pictures and words. *Nature*, 253(5491):437.

77. Potter MC, Kroll JF, Yachzel B, Carpenter E, Sherman J (1986) Pictures in sentences: understanding without words. *J. Exp. Psychol. Gen.* 115(3):281.

78. Devlin J, Chang MW, Lee K, Toutanova K (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

79. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.

80. Lucy L, Gauthier J (2017) Are distributional representations ready for the real world? Evaluating word vectors for grounded perceptual meaning. *arXiv preprint arXiv:1705.11168*.

81. Willits JA, Amato MS, MacDonald MC (2015) Language knowledge and event knowledge in language use. *Cogn. Psychol*. 78:1-27.

82. Bright P, Moss H, Tyler LK (2004) Unitary vs multiple semantics: PET studies of word and picture processing. *Brain Lang.* 89(3):417-32.

83. Devereux BJ, Clarke A, Marouchos A, Tyler LK (2013) Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *J. Neurosci.* 33(48):18906-16.

84. Fairhall SL, Caramazza A (2013) Brain regions that represent amodal conceptual knowledge. *J. Neurosci.* 33(25):10552-8.

85. Handjaras G, Leo A, Cecchetti L, Papale P, Lenci A, Marotta G, Pietrini P, Ricciardi E (2017) Modality-independent encoding of individual concepts in the left parietal cortex. *Neuropsychologia* 105:39-49.

86. Sevostianov A, Horwitz B, Nechaev V, Williams R, Fromm S, Braun AR (2002) fMRI study comparing names versus pictures of objects. *Hum. Brain Mapp.* 16(3):168-75.

87. Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RS (1996) Functional anatomy of a common semantic system for words and pictures. *Nature* 383(6597):254.

88. Visser M, Jefferies E, Embleton KV, Lambon Ralph MA (2012) Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J. Cogn. Neurosci.* 24(8):1766-78.

89. Wagner AD, Desmond JE, Demb JB, Glover GH, Gabrieli JD (1997) Semantic repetition priming for verbal and pictorial knowledge: A functional MRI study of left inferior prefrontal cortex. *J. Cogn. Neurosci.* 9(6):714-26.

90. Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9(1):97-113.

91. Mahowald K, Fedorenko E (2016) Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. *NeuroImage* 139:74-93.

92. Scott TL, Gallée J, Fedorenko E (2017) A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cogn. Neurosci.* 8(3):167-76.

93. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67(1):1-48.

94. Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* 82(13):1–26.

95. Pujol J, Deus J, Losilla JM, Capdevila A (1999) Cerebral lateralization of language in normal left-handed people studied by functional MRI. *Neurology* 52(5):1038-1038.

96. Howard D, Patterson K (1992) The Pyramids and Palm Trees Test: A test of semantic access from words and pictures. *Thames Valley Test Company*.

97. Forster KI, Forster JC (2003) DMDX: A Windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35(1):116-24.

98. Woolgar A, Duncan J, Manes F, Fedorenko E (2018) Fluid intelligence is supported by the multiple-demand system not the language system. *Nat. Hum. Behav.* 2(3):200.

Supplemental Information

Experiment 1

Whole-brain random effects analysis

The analysis was conducted by using the spm_ss toolbox (available at http://www.nitrc.org/projects/spm_ss), which interfaces with SPM and the CONN toolbox (https://www.nitrc.org/projects/conn). The results were thresholded at p=0.001, and resulting clusters were FDR-corrected at p=0.05.

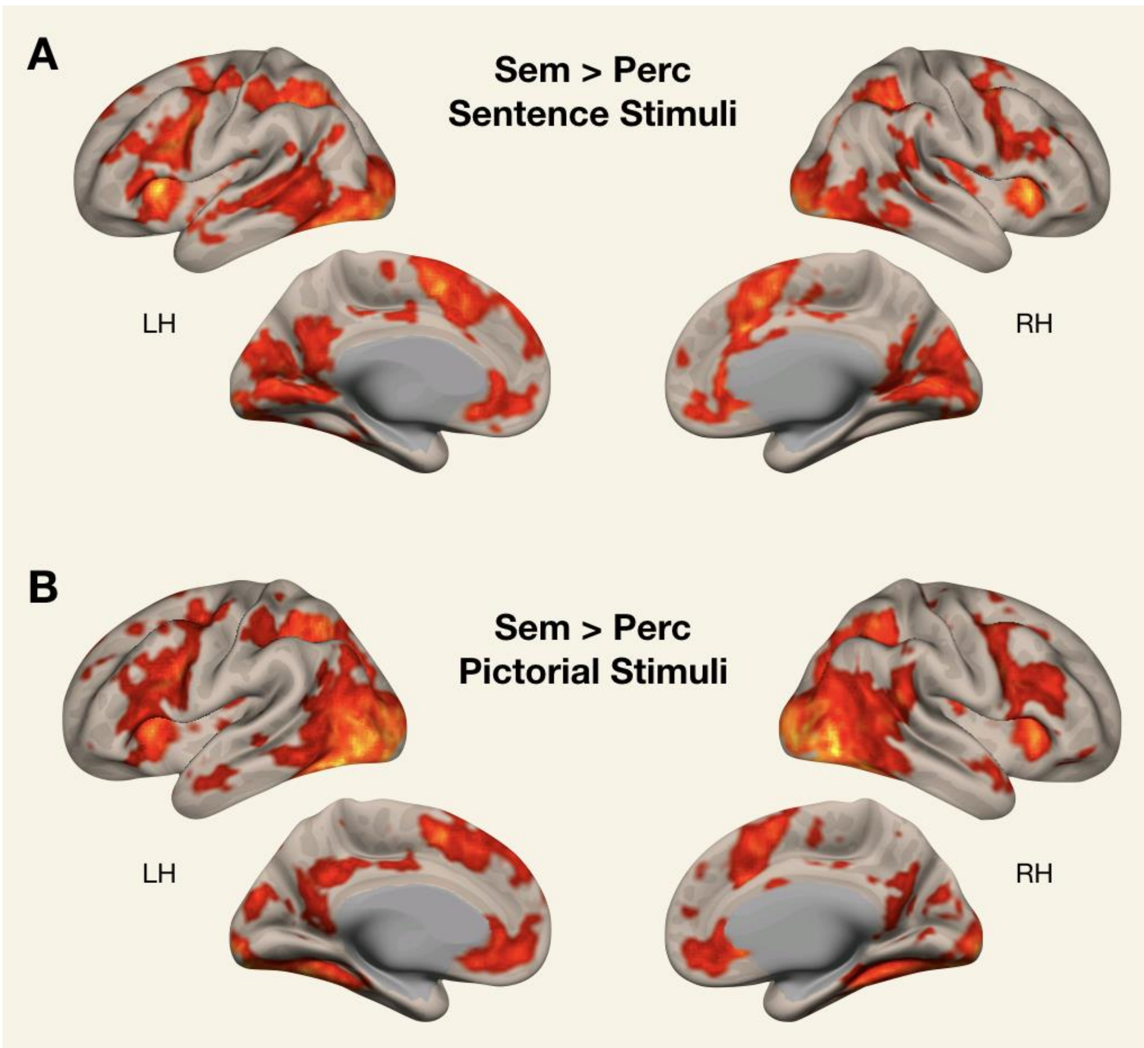

Figure S1. *Whole-brain random effects group analysis for Semantic > Perceptual task contrast, conducted separately on the data from sentence trials (a) and picture trials (b).*

Statistical analyses

The formula used for the mixed linear effects model was *EffectSize ~ StimType * Task + (1 | ROI) + (1 | Subject)*. The formula used for the follow-up comparisons was *EffectSize ~ StimType * Task + (1 | Subject)*.

Behavioral Results

Average response rate was 92% and did not vary significantly across tasks. Average response times were 1.27 s (SD = 0.46) for the semantic sentence task, 1.16 s (SD = 0.38) for the perceptual sentence task, 1.22 s (SD = 0.35) for the semantic picture task, and 1.19 (SD = 0.36) for the perceptual picture task. Average accuracies were 0.81 for the semantic sentence task, 0.79 for the perceptual sentence task, 0.75 for the semantic picture task, and 0.75 for the perceptual picture task. Note that, due to a technical error, behavioral data from 14 out of 21 participants was only recorded for one of the two runs.

Note on the SPM analysis software

We used SPM5 for our analyses. Note that SPM was only used for preprocessing and basic data modeling. These aspects of SPM's functionality have not changed much between the different versions. The reason for the use of an older version is because we have projects that use data collected over the last 15 years, and we want to keep preprocessing and first-level analysis the same across the hundreds of subjects in our database, which are pooled across analyses for many projects. We have directly compared the outputs of data preprocessed and modeled in SPM5 vs. e.g., SPM12, and the outputs are nearly identical.

Experiment 2

Ethics approval was granted by the UCL Research Ethics Committee (LC/2013/05). All participants provided informed consent prior to taking part in the study.

Participant information

Two patients with global aphasia, S.A. and P.R., participated in the study. Both participants were male and native English speakers. S.A. was 22 years 5 months post-onset of his neurological condition, and P.R. was 14 years 7 months post-onset. S.A. had a subdural empyema in the left sylvian fissure, with associated meningitis that led to a secondary vascular lesion in left middle cerebral artery territory. P.R. also had a vascular lesion in left middle cerebral artery territory.

Both patients were classified as severely agrammatic (Table S1). Whereas they had some residual lexical comprehension ability, scoring well on tasks involving word-picture matching and synonym matching across spoken and written modalities, their lexical production was impaired. Both patients failed to correctly name a single item in a spoken picture-naming task. S.A. displayed some residual word production ability, scoring 24 out of 60 in a written picture-naming task. P.R., however, performed poorly in the written task, correctly naming just 2 out of 60 items. S.A. and P.R.'s syntactic processing was also severely disrupted. They scored near or below chance on both active and passive forms in sentence comprehension tasks that involved pairing a spoken or written sentence with a corresponding picture in the presence of distractors, in which protagonists' roles were reversed. They also obtained scores around chance in written grammaticality judgment assessments, which required them to determine whether a written sentence was grammatical or not. Phonological working memory was evaluated by means of a digit span test, using a recognition paradigm that did not require language production. Both participants had a digit span that was sufficiently long to process the types of sentences in the syntactic assessments, and thus the aforementioned difficulties could not be attributed to phonological working memory problems.

As the semantic task involved visual event processing, the patients' performance in non-verbal reasoning tasks was also evaluated (Table 2). Both participants performed well on Pyramids and Palm Trees Test (PPT 3-picture version, Howard & Patterson, 1992), a commonly used test that probes people's ability to form conceptual associations, and on Raven's Colored Matrices, (Raven, Raven, & Court, 1998a), a test where colored patterns are missing an element, which has to be selected from among several alternatives. S.A. also scored highly (53/60) on Raven's Progressive Matrices (Raven, Raven, & Court, 1998b), a test that is similar to the Colored Matrices, but that includes a broader, more abstract range of increasingly difficult patterns that require participants to hold multiple variables in working memory, and place a greater demand on response inhibition and analytic capacity. P.R. did not perform quite as well on the Progressive Matrices, scoring 36 out of 60. Finally, both participants undertook the Visual Patterns Test (VPT; Della Sala, Gray, Baddeley, & Wilson, 1997), which assesses visual working memory by getting participants to memorize a series of black and white squares arranged to form different patterns of varying sizes and increasing complexity. Whereas S.A. placed in the 90th percentile (in relation to adults in the same age range with no neurological impairment), P.R. scored in the 40th percentile.

Each patient performed both the semantic task and the sentence-picture matching task, with a 7-months period between the two.

Table S1. Results of linguistic assessments for participants with global aphasia.

Lexical Tests	**Chance Score**	**S.A.**	**P.R.**
ADA spoken word picture matching	16.5	60/66*	61/66*
ADA written word picture matching	16.5	62/66*	66/66*
ADA spoken synonym matching	80	123/160*	113/160*
ADA written synonym matching	80	121/160*	145/160*
PALPA 54 spoken picture naming	n/a	0/60*	0/60*
PALPA 54 written picture naming	n/a	24/60	2/60
Syntactic Tests			
Comprehension of spoken reversible sentences	50	49/100	38/100
Comprehension of written reversible sentences	50	42/100	49/100
Written grammaticality judgments	20	26/40*	21/40
Verbal Working Memory			
PALPA 13 digit span (recognition)	n/a	3 items	4 items

* indicates above chance performance ($p < .05$)

The tests were from the Action for Dysphasic Adults (ADA) Auditory Comprehension Battery (Franklin, Turner, & Ellis, 1992) and the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992), or were developed for the purpose of the study.

Table S2. Results of non-linguistic assessments for participants with global aphasia.

Reasoning Tests	**S.A.**	**P.R.**
Ravens Colored Matrices	36/36	34/36
Ravens Progressive Matrices	53/60	36/60
Pyramids and Palm Trees (3 picture version)	50/52	47/52
Visual Pattern Test (age group percentile)	11.5 (90^{th} percentile)	8.6 (40^{th} percentile)

References

Franklin, S., Turner, J. E., & Ellis, A. W. (1992). *The ADA auditory comprehension battery*. York, U.K.: University of York.

Howard, D., & Patterson, K. E. (1992). *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Thames Valley Test Company.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessment of language processing in aphasia.* Hove, U.K.: Psychology Press.