Supplementary Material

Appendix A. Experimental design

Material

Participants read chapter 9 of *Harry Potter and the Sorcerer's Stone* [16]. We chose this chapter because it involves many characters and spans multiple locations and scenes. We chose a famous book series because we hypothesized all subjects already had characteristic mental representations of the different characters and locations, and that at least a part of this representation would remain constant throughout the reading of chapter 9. This assumption allows us to use data from the entire chapter to look for the representation of the different characters, e.g. the protagonist Harry Potter. In contrast, had we chosen an unfamiliar story in which we learn about the protagonist's personality throughout the text, the mental representation of this protagonist will arguably change more than Harry's would.

Participants

fMRI data was collected from 9 subjects (5 females and 4 males) recruited through Carnegie Mellon University, aged 18 to 40 years. The participants were all native English speakers and right handed. They were chosen to be familiar with the material: we made sure they had read the Harry Potter books or seen the movie series and were familiar with the characters and the story. All the participants were screened for safety, signed the consent form and were compensated for their participation. Data from one of the subjects was excluded from the analysis because of an artifact that was not removed by our preprocessing procedure.

Design

The words of the story were presented in rapid serial visual format [30]. Words were presented one by one at the center of the screen for 0.5 seconds each (see Fig. 5). The background was gray and the font was black. We used MATLAB and the Psychophysics Toolbox extensions [31–33].

The chapter was divided into four runs, of approximately 11 minutes each. Subjects had short breaks between runs. Each run started with a fixation period of 20 seconds in which the subjects stared at a cross in the middle of the screen. The words presentation started after the fixation period. The total length of the runs was 45 minutes, during which about 5200 words were presented. Chapter 9 was presented in its entirety without modifications and each subject read the chapter only once.

Before the experiment, we supplied the subjects with a summary of the events preceding chapter 9 and a summary of the main characters and concepts in *Harry Potter and the Sorcerer's Stone* to refresh their memory. We also instructed them to practice rapid serial presentation by viewing a video that replicated the parameters of our design, but with another story (*The Tale of Peter Rabbit* [34]). On the day of the experiment, the subjects were instructed to lay in the scanner and read the chapter as naturally as possible while remaining alert.

fMRI procedure

Functional images were acquired on a Siemens Verio 3.0T scanner (Siemens, Erlangen, Germany) at the Scientific Imaging & Brain Imaging Center at Carnegie Mellon University, using a T2* sensitive echo planar imaging pulse sequence with repetition time (TR)=2s, echo time=29 ms, flip angle=79°, 36 slices and $3 \times 3 \times 3$mm voxels. Anatomical volumes were acquired with a T1-weighted 3D-MPRAGE pulse sequence.

Data preprocessing

We used the MATLAB suite SPM8 [35] to preprocess the data. Each subject's functional data underwent realignment, slice timing correction and co-registration with the subject's anatomical scan, which was segmented into grey and white matter and cerebro-spinal fluid. The subject's scans were normalized to the Montreal Neurological Institute (MNI) space and smoothed with a 6 × 6 × 6mm Gaussian kernel smoother.

Using the Python toolbox PyMVPA [36], we masked the functional data using the segmented anatomical mask, discarding cerebrospinal-fluid voxels. The data was then detrended in MATLAB by running a high-pass filter with a cut-off frequency of 0.005Hz. Visual inspection of the time course of a large number of voxels showed that this threshold was enough to get rid of large block effects and slow trends in the data.

Finally, we selected voxels from each subject, keeping only voxels in 78 cortical Regions Of Interest (ROIs), defined using the AAL brain atlas [37], excluding the cerebellum and white matter. We ended up with an average of 29227 voxels per subject. The anatomical union (number of MNI voxel locations for which at least one subject had a voxel) of these 6 subject's brains was a set of 41073 voxel locations.

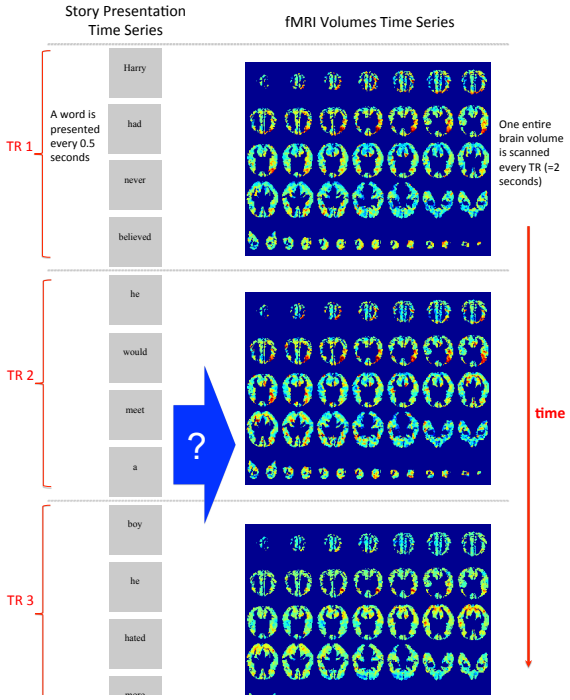

Figure 5. Illustration of our fMRI experimental protocol. Words from a story are presented serially for 0.5 seconds each while recording brain activity with fMRI at a rate of one entire brain image each 2 seconds. Our goal is to model how fMRI neural activity during reading reflects the perceptual and conceptual features of the story. Each fMRI activity volume is shown here in 36 horizontal slices. Going right to left through the slices, then bottom-up, corresponds to looking at slices from the bottom of the brain up. Within each slice, the top of the slice corresponds to the posterior of the brain, and the right side of the slice corresponds to the left side of the brain. The images are on a scale from blue to red where blue indicates negative deviation from baseline and red indicates positive deviations. A TR is the time needed to record one brain volume, and is 2 seconds in our experiment.