Psychological Science

Brain Mechanisms Underlying Reality Monitoring for Heard and Imagined Words

Eriko Sugimori, Karen J. Mitchell, Carol L. Raye, Erich J. Greene and Marcia K. Johnson

Brain Mechanisms Underlying Reality Monitoring for Heard and Imagined Words

**Eriko Sugimori, Karen J. Mitchell, Carol L. Raye,
Erich J. Greene, and Marcia K. Johnson**
Yale University

Abstract

Using functional MRI, we investigated reality monitoring for auditory information. During scanning, healthy young adults heard words in another person's voice and imagined hearing other words in that same voice. Later, outside the scanner, participants judged words as "heard," "imagined," or "new." An area of left middle frontal gyrus (Brodmann's area, or BA, 6) was more active at encoding for imagined items subsequently correctly called "imagined" than for items incorrectly called "heard." An area of left inferior frontal gyrus (BA 45, 44) was more active at encoding for items subsequently called "heard" than "imagined," regardless of the actual source of the item. Scores on an Auditory Hallucination Experience Scale were positively related to activity in superior temporal gyrus (BA 22) for imagined words incorrectly called "heard." We suggest that activity in these areas reflects cognitive operations information (middle frontal gyrus) and semantic and perceptual detail (inferior frontal gyrus and superior temporal gyrus, respectively) used to make reality-monitoring attributions.

Source monitoring refers to mechanisms by which people make attributions about the origin of mental experiences. According to the source-monitoring framework (SMF; Johnson, Hashtroudi, & Lindsay, 1993), this includes evaluating qualitative characteristics such as perceptual, semantic, temporal, and spatial detail and information about the cognitive operations engaged during encoding. In the present study, we were concerned with one class of such attributions, *reality monitoring*—discriminating perceived from self-generated events (e.g., differentiating a memory of a previous perception from that of a previous imagination; Johnson & Raye, 1981).

Reality monitoring is an essential but imperfect cognitive function (Johnson & Raye, 1981). External misattributions of self-generated information (i.e., reality-monitoring errors) are fairly common, and much has been learned from systematically studying such errors in the laboratory (see Johnson, 2006; Johnson, Raye, Mitchell, & Ankudowich, 2011; Lindsay, 2008, for reviews). For example, good imagers make more misattributions of self-generated information to an external source than do poor imagers, presumably because good imagers create representations that are rich in visual and other contextual details, which makes them similar to perceptions, or they do so easily or spontaneously, which leaves few records of cognitive effort (Dobson & Markham, 1993; Johnson, Raye, Wang, & Taylor, 1979). Records of the cognitive operations (e.g., the act of imagining) carried out at the time of encoding can be used during remembering as a cue that the information was self-generated. Anything that decreases production of, or attention to, cognitive operations information is likely to increase errors (Finke, Johnson, & Shyi, 1988). Hence, it is likely that both more vivid perceptual information and less vivid information about cognitive operations play a role in misattributing self-generated information to perception.

The idea that reality-monitoring errors are based, at least in part, on self-generated perceptual information is supported by recent functional MRI (fMRI) studies. Gonsalves et al. (2004) scanned participants as they saw

Corresponding Author:
Marcia K. Johnson, Department of Psychology, Yale University, P. O. Box 208205, New Haven, CT 06520-8205
E-mail: marcia.johnson@yale.edu

pictures and imagined pictures. A subsequent memory analysis identified an area of precuneus—a brain region associated with visual processing—where activity was greater for imagined items that participants subsequently incorrectly called "seen" than for those items correctly called "imagined" (see also, e.g., Kensinger & Schacter, 2005). In addition, activity in posterior visual areas during test is associated with attributions to perception, regardless of the actual source (i.e., seen vs. imagined pictures; Kensinger & Schacter, 2006). In contrast, activity in left lateral and medial prefrontal cortex (PFC) at both encoding (e.g., Kensinger & Schacter, 2005) and test (e.g., Dobbins & Wagner, 2005; Kensinger & Schacter, 2006; Mitchell et al., 2008; Simons, Gilbert, Owen, Fletcher, & Burgess, 2005) is associated with correct attributions of self-generated information. Presumably, these PFC areas process or represent either the cognitive operations engaged during encoding (e.g., those involved in imagining or in carrying out a particular encoding task, such as estimating cost) or the conceptual-semantic output of these processes.

Most studies of reality monitoring in healthy individuals have used visual information, but reality-monitoring errors occur for auditory information as well (Foley, Johnson, & Raye, 1983; Johnson, Foley, & Leach, 1988). For example, Johnson et al. (1988) found that participants were more likely to misattribute imagined words to a speaker if they had imagined the words in the speaker's voice than if they had imagined them in their own voice. Evidence from fMRI regarding the specific mechanisms of misattributions of auditory information in healthy individuals, however, is missing. Hence, the primary aim of the current study was to specify brain mechanisms underlying external misattributions of auditory information in healthy individuals.

In the scanner, healthy young adults heard some words in another person's voice and imagined hearing other words in that person's voice. We expected that increased activity in brain areas associated with hearing (primary and secondary auditory cortex) should be associated with heard items, and increased activity in areas associated with cognitive operations or the output of those operations such as semantic information (e.g., medial and left lateral prefrontal cortex) should be associated with imagined items. Later, we presented participants with visual words and asked them whether they heard each word, they imagined it, or it was new: Our primary interest was in identifying which of the areas with differential encoding activity predicted subsequent reality-monitoring performance. We expected that activity in areas associated with cognitive operations would be related to correct attributions of imagined items. We also expected, parallel to the findings of Gonsalves et al. (2004), that misattributions of imagined words as heard would be related to activity in areas associated with representing the output of such imagining (auditory or semantic detail).

Although most studies have investigated reality monitoring in the domain of memory, similar factors and mechanisms likely contribute to reality monitoring in the domains of perception and belief (Johnson, 1988). That is, the SMF is relevant for understanding false perceptions and beliefs (e.g., hallucinations and delusions; see Woodward & Menon, 2013, for a review and discussion) as well as false memories (Johnson et al., 2011). Consistent with this idea, studies have found that nonclinical participants who are prone to auditory hallucinations show increased external misattributions of self-generated information in various reality-monitoring tasks, relative to participants who are not prone to hallucinations (e.g., Larøi, Van der Linden, & Marczewski, 2004; Sugimori, Asai, & Tanno, 2011). Thus, we also assessed the relation between participants' proneness to auditory hallucinations and brain activity associated with misattributions of imagined auditory information. This should help further refine the understanding of the overlap in mechanisms underlying reality-monitoring errors and on-line misperceptions (e.g., poor cognitive operations information, overly vivid auditory information). Given evidence from the clinical literature about the importance of superior temporal gyrus (STG) in auditory hallucinations in patients (Kompus, Westerhausen, & Hugdahl, 2011), we expected a relation between proneness to auditory hallucinations and STG activity.

Method

Participants

Twenty participants (11 females, 9 males; all right-handed; mean age = 22.2 years, *SD* = 3.8, range = 18–29) self-reported being in good health, with no history of stroke, heart disease, primary degenerative neurological disorder, or psychiatric diagnosis; they had normal or corrected-to-normal vision and were not taking psychotropic medications. All participants were paid. The Human Investigation Committee of Yale University Medical School approved the protocol; informed consent was obtained from all participants.

Design

There were two phases in the study. Phase 1 (encoding) was carried out while participants were scanned. There were three types of trials: heard words, imagined words, and shapes. Phase 2 (test) was a surprise source-memory test conducted outside the scanner; participants responded "heard," "imagined," or "new" to heard, imagined, and new items, respectively.

Materials

A total of 308 words were used: 10 for practice outside the scanner, 10 for practice inside the scanner, and 288 for experimental trials. The experimental words were assigned to three 96-word lists equated on several characteristics, such as frequency, concreteness, familiarity, imageability, meaningfulness, and number of letters, phonemes, and syllables, using the MRC Psycholinguistic Database (Coltheart, 1981). The three lists were rotated across the heard, imagined, and new conditions. A recorded male voice read aloud a news story, instructions for Phase 1, and the words for the heard condition. During encoding, we also presented 30 nonauditory shape trials: figures composed of four contiguous shapes (with various numbers of circular and square shapes).

We used an Auditory Hallucination Experience Scale (AHES) to measure auditory verbal-hallucination proneness. It had 12 items adapted from the Launay Slade Hallucination Scale (Launay & Slade, 1981), a widely used instrument for measuring disposition to hallucinate. One statement measuring predisposition to visual hallucination ("On occasion, I have seen a person's face in front of me when no one was in fact there") was deleted; "I have heard the voice of the Devil" and "In the past, I have heard the voice of God speaking to me" were replaced with "I have heard a voice inside me calling me toward bad things" and "I have heard a voice inside me telling me to do good things," respectively.

Procedure

Task instructions, read in the same voice as the words in the heard condition, were given outside the scanner via headphones. Participants were told this was a study comparing imagination with perception. To ensure participants were familiar with the sound of the voice they would later be asked to imagine, we also played them a recording of the voice reading a news story for approximately 3.5 min.

On each trial in Phase 1, participants first saw a cue in the middle of the computer screen for 2 s that told them whether they would hear a word (heard condition), they should imagine hearing a word in the pre-exposed "other" voice (imagined condition), or they would see a shape (shape condition). The cues "hear" and "imagine" were each followed by a new word presented visually for 2 s, and concurrently, participants either heard the male voice saying the word or imagined his voice saying that word. When the word disappeared from the screen, participants were cued to rate, by pressing buttons with the first three fingers of their right hand, how well they heard or imagined hearing the word (1 = *low quality*, 2 = *average*, 3 = *high quality*). Participants had 2 s to make their response.

The "shape" cue was followed by a figure composed of four adjoined circular and square shapes presented visually for 2 s. When the figure disappeared from the screen, participants were cued to estimate, using the same three buttons, whether there were more square or circular shapes (1 = *more square shapes*, 2 = *an equal number of square and circular shapes*, 3 = *more circular shapes*). Participants had 2 s to make their response.

Between trials, there was an unfilled 6-s intertrial interval to allow the hemodynamic response to return to baseline. Participants practiced five heard trials, five imagined trials, and three shape trials in pseudorandom order before getting into the scanner and practiced the same number of trials again (with different items) during the structural scans so that the volume of the headphones could be adjusted as needed for each participant. Phase 1 consisted of six runs in the scanner. In each run, participants heard 16 words, imagined the voice saying 16 words, and saw 5 shapes; these trials were pseudorandomly intermixed.

About 5 min after the scan, there was a surprise source-memory test in a separate room. This test consisted of 96 heard items, 96 imagined items, and 96 new items (which had not been presented elsewhere in the experiment) sequentially presented visually on a computer screen for 4 s each in pseudorandom order. Participants indicated, for each word, if they had heard it, if they had imagined it, or if it was new, pressing the keys 1, 2, and 3, respectively, using the first three fingers of their right hand.

Imaging details

Images were acquired using a 3T Siemens Trio scanner. After anatomical localizer scans, functional images were acquired with a single-shot echoplanar gradient-echo pulse sequence (repetition time = 2,000 ms, echo time = 25 ms, flip angle = 80°, field of view = 240 mm). The 36 oblique axial slices were 3.5 mm thick with an in-plane resolution of 3.75 × 3.75 mm; they were aligned with the anterior commissure-posterior commissure line. Each run began with 12 s of blank screen to allow tissue to reach steady-state magnetization and was followed by a 1-min rest interval.

fMRI analyses

After reconstruction, time series were shifted by sinc interpolation to correct for slice-acquisition times. Data were motion-corrected using a 6-parameter automated algorithm (Automated Image Registration, or AIR; Woods, Cherry, & Mazziotta, 1992). A 12-parameter AIR algorithm was used to coregister participants' images to a common reference brain. Data were mean-normalized across time and participant and spatially smoothed (3-D, 8-mm, full-width half-maximum Gaussian kernel). We used

voxel-based analysis of variance (ANOVA) with participant as a random factor and all other factors fixed (NeuroImaging Software; Laboratory for Clinical Cognitive Neuroscience, University of Pittsburgh, and Center for the Study of Brain, Mind, and Behavior, Princeton University). *F* maps were transformed to Talairach space using Analysis of Functional NeuroImages (AFNI) software (http://afni.nimh.nih.gov/afni), and areas of activation were localized using AFNI and Talairach Daemon software (Lancaster et al., 1997, 2000) and were visually checked against atlases.

Because we were interested in event-related responses, we first identified regions showing an Encoding Condition (heard words, imagined words, shapes)[1] × Time Within Trial (Scans 1–6) interaction with a minimum of 10 contiguous voxels, each significant at $p < 1.0 \times 10^{-13}$ (Lieberman & Cunningham, 2009). For the regions thus identified, subsequent memory analyses were conducted on the mean percentage change (from Time 1) at Times 3 and 4 or Times 4 and 5 (depending on maximal differences in activity per region), using a 2 (encoding condition: heard words, imagined words) × 3 (response at test: "heard," "imagined," "new") ANOVA. We also calculated, for these same regions, Pearson's correlation coefficients between these mean percentage change scores for each participant for each combination of encoding condition and subsequent response, and scores on the AHES.

Results

Behavioral data

On-line ratings (maximum = 3.00) indicated that participants were able to hear and imagine the words well. Participants gave higher ratings to heard items ($M = 2.71$) than to imagined items ($M = 2.50$), $F(1, 19) = 15.57$, $p = .001$. They also gave higher ratings to items they subsequently called "heard" ($M = 2.68$) than to those they called "imagined" ($M = 2.59$) or "new" ($M = 2.56$), $F(2, 38) = 7.76$, $p = .001$; the ratings of items called "imagined" did not differ significantly from those called "new."

Figure 1 shows mean proportion of each response type during the source-memory test for heard, imagined, and new items. For old words misattributed as new, there was no significant difference between the proportion of items that had been encoded in the heard condition ($M = 0.43$) versus the imagined condition ($M = 0.47$), $F(1, 19) = 0.36$, $p = .55$, $\eta_p^2 = .019$. There was also no significant difference in the proportion of new words misattributed as heard ($M = 0.14$) versus imagined ($M = 0.13$), $F(1, 19) = 0.11$, $p = .74$, $\eta_p^2 = .006$. Thus old/new recognition was approximately the same for heard and imagined items.

Of central interest, for old items, a 2 (item type: heard words, imagined words) × 2 (response: "heard," "imagined") ANOVA was conducted. There was a significant
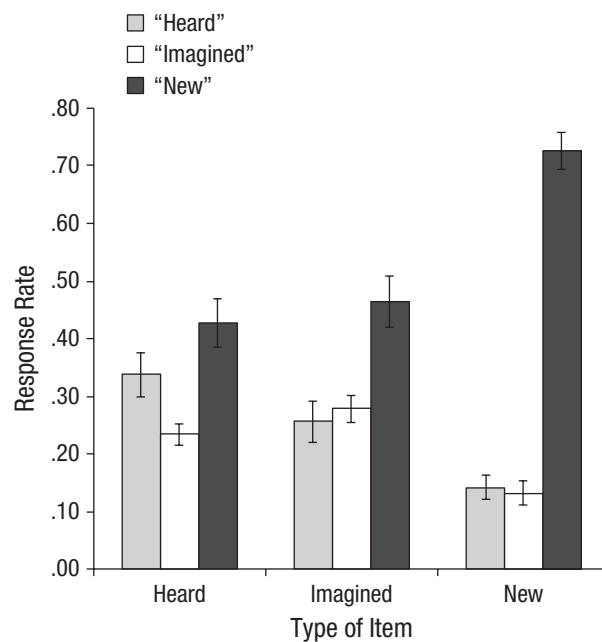

Fig. 1. Mean response rate at test as a function of item type and participants' response. Error bars show standard errors of the mean.

interaction between item type and response, $F(1, 19) = 7.33$, $p = .01$, $\eta_p^2 = .278$. For heard items, participants made a higher proportion of correct "heard" responses ($M = 0.34$) than incorrect "imagined" responses ($M = 0.23$), $F(1, 38) = 5.84$, $p = .02$, $\eta_p^2 = .133$. For imagined items, there was no significant difference between the proportion of correct "imagined" responses ($M = 0.28$) and incorrect "heard" responses ($M = 0.26$; $p = .60$).[2] This pattern is remarkably similar to the pattern seen in Johnson et al. (1988, Experiment 2), in which old/new recognition was much better (hits ~75% did not differ across the heard and imagined conditions), which indicates that reality-monitoring difficulty does not depend on low levels of recognition.

From the behavioral data, it might be tempting to conclude that participants had some source information about heard words but none about imagined words and hence simply guessed the source on the imagined items recognized as old. However, the fMRI data provide evidence that participants were not simply guessing on the imagined items but rather, as predicted by the SMF, were responding based on qualities of their subjective experience.

fMRI data

Table 1 shows all areas identified in our initial whole-brain analysis. The section headings in the table indicate the main effects of encoding condition identified by the subsequent memory analyses. Of primary interest were areas that showed a significant main effect of response

Table 1. Regions of Activation: All Areas Identified as Showing an Encoding Condition × Time Within Trial Interaction in an Initial Whole-Brain Analysis of Variance

Hemisphere	Brodmann's area	Anatomical area	Talairach coordinates			Maximum F for the region		Effect of encoding condition in subsequent analyses	
			x	y	z	$F(10, 190)$	η_p^2	$F(1, 19)^a$	η_p^2
		Imagined words > heard words							
Left	**45, 44**	**Inferior frontal gyrus**	**–47**	**15**	**19**	**24.16**	**.560**	**21.96**	**.536**
Left	**6, (4)**	**Middle frontal gyrus (precentral gyrus)**	**–29**	**–10**	**48**	**13.67**	**.418**	**15.81**	**.454**
Right	6	Middle frontal gyrus (superior frontal gyrus)	23	–5	50	18.87	.498	6.55	.256
Right	6	Precentral gyrus	46	2	28	19.28	.503	4.76	.200
Medial	31, 19	Precuneus, cingulate gyrus	–8	–58	29	15.71	.453	2.99	.136
Left	19, 18, 7	Middle occipital gyrus, superior occipital gyrus, fusiform gyrus, inferior parietal lobule	–31	–73	24	39.23	.674	6.53	.256
Right	7, 19, 37, 39, 40	Precuneus, middle occipital gyrus, superior occipital gyrus, fusiform gyrus, supramarginal gyrus, angular gyrus, inferior parietal lobule, superior parietal lobule	21	–66	38	67.33	.780	8.79	.316
		Heard words > imagined words							
Left	**22, 21, 39, 37**	**Superior temporal gyrus, middle temporal gyrus, tranverse temporal gyrus, inferior parietal lobule**	**–46**	**–22**	**6**	**64.33**	**.772**	**139.05**	**.880**
Right	**22, 21, 39, 37**	**Superior temporal gyrus, middle temporal gyrus, tranverse temporal gyrus, inferior parietal lobule**	**46**	**–23**	**57**	**71.63**	**.790**	**145.45**	**.884**

Note: Bolded areas are discussed in the text. Talairach coordinates are shown for the voxel with the maximum F value in each area of activation. For identified areas, subsequent memory analyses were conducted on the mean percentage change (from Time 1) at Times 3 through 4 or Times 4 through 5 (depending on maximal differences in activity), using a 2 (encoding condition: heard words, imagined words) × 3 (response type at test: "heard," "imagined," "new") analysis of variance. The main effects of encoding condition for these areas are indicated in the table's section headings, along with the F values in the rightmost column. Main effects of response type at test and interactions are reported in the text. For each area of activation, the major anatomical regions and Brodmann's area numbers are listed starting with the location of the local maxima and extending from there (parentheses indicate a small extent relative to other areas listed).
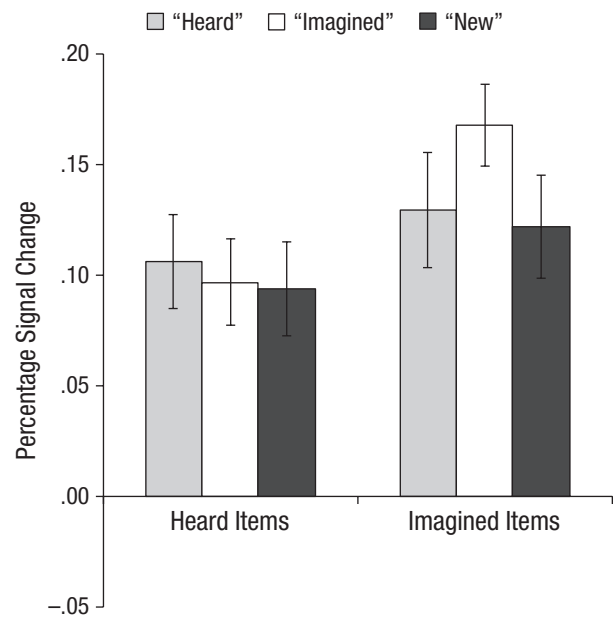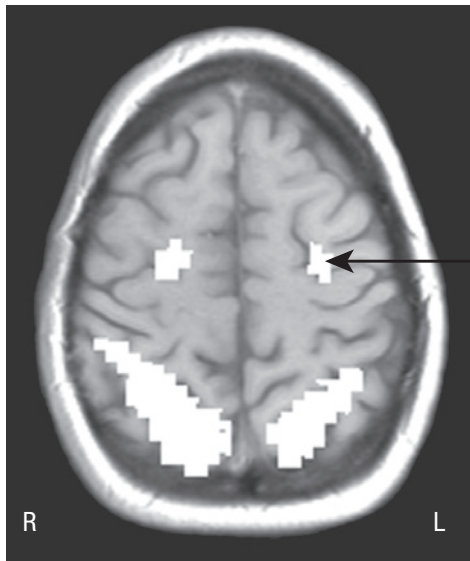[a]ps < .05.

type at test or an Encoding Condition × Response Type interaction.

An area including left middle frontal gyrus (MFG), extending into precentral gyrus—Brodmann's area (BA) 6, (4) showed a significant interaction between encoding condition and response type at Times 3 through 4, $F(2, 38) = 3.36$, $p = .04$, $\eta_p^2 = .150$ (Fig. 2a). For heard items, there was no significant difference in activity among

responses ($F < 1$), but for imagined items, $F(2, 38) = 4.05$, $p = .03$, $\eta_p^2 = .176$, there was more activity for items subsequently correctly called "imagined" than for items incorrectly called "heard," $F(1, 19) = 4.88$, $p = .03$, $\eta_p^2 = .204$, or "new," $F(1, 19) = 7.06$, $p = .02$, $\eta_p^2 = .271$; "heard" and "new" responses did not differ significantly ($p > .05$).

An area including left inferior frontal gyrus (IFG; BA 45, 44; Fig. 2b) showed a significant main effect of
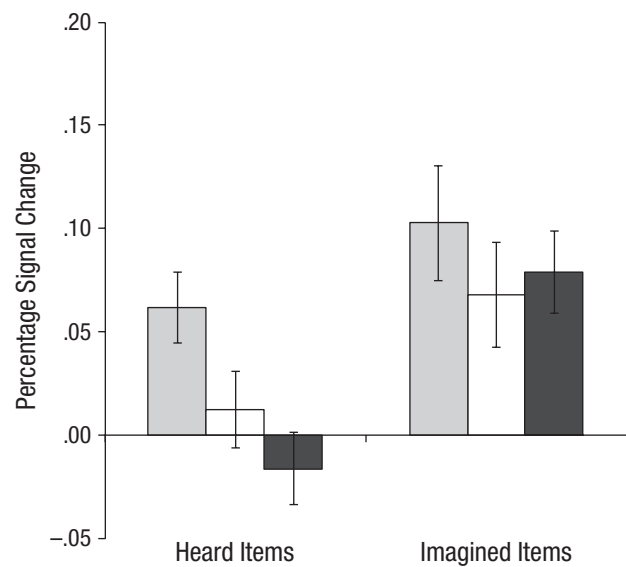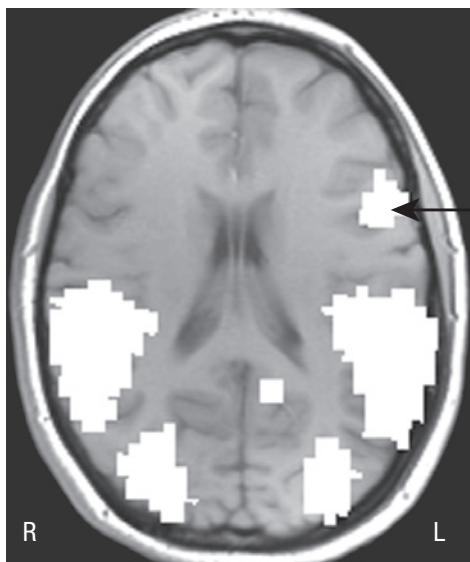
a

b

Fig. 2. Results for areas of prefrontal cortex where encoding activity differentially predicted subsequent reality-monitoring judgments: an area of (a) left middle frontal gyrus (MFG) extending into precentral gyrus (Brodmann's area 6, (4)) and (b) left inferior frontal gyrus (IFG; Brodmann's area 45, 44). The graphs show mean percentage change in blood-oxygen-level-dependent (BOLD) signal at encoding for heard and imagined items as a function of participants' response at test. Error bars show standard errors of the mean. R = right, L = left.

response type at Times 4 through 5, $F(2, 38) = 3.62$, $p = .04$, $\eta_p^2 = .167$: "Heard" responses were greater than "imagined" responses, $F(1, 19) = 6.34$, $p = .04$, $\eta_p^2 = .251$, or "new" responses, $F(1, 19) = 4.32$, $p = .02$, $\eta_p^2 = .185$; "imagined" and "new" responses were not significantly different ($p > .05$). The Encoding Condition × Response Type interaction was not significant ($Fs < 1$).

We were also interested in areas that showed correlations with the AHES. There were only two such areas: Large bilateral areas of STG that extended to include

middle and transverse temporal gyri and inferior parietal lobule (Figs. 3a and 3b) showed only main effects of encoding condition at Times 4 through 5 (heard > imagined; see Table 1). No other subsequent memory effects were significant. As shown in Figures 3c and 3d, activity in these regions for imagined items called "heard" was positively related to scores on the AHES (right STG: $r = .54$, $p = .01$; left STG: $r = .46$, $p = .04$). People more prone to auditory hallucinations showed more activation in this region for imagined items later mistakenly called "heard."

Discussion

Consistent with expectations that frontal regions would reflect engaging cognitive operations, our results showed that at encoding, there was greater activity for imagined than heard words in left MFG and left IFG. Consistent with the expectation that temporal regions would reflect processing auditory information, areas including bilateral STG were more active for heard than imagined words. Our main question was how brain activity in these regions during encoding was related to subsequent reality-monitoring judgments for whether words were heard or imagined.

Greater activity in left MFG (BA 6, [4]) during encoding for imagined items was associated with later correctly calling the items "imagined" rather than incorrectly calling them "heard." This pattern suggests that records of encoding activity in this area as participants imagined the words functioned later as cues at test. Similar areas are
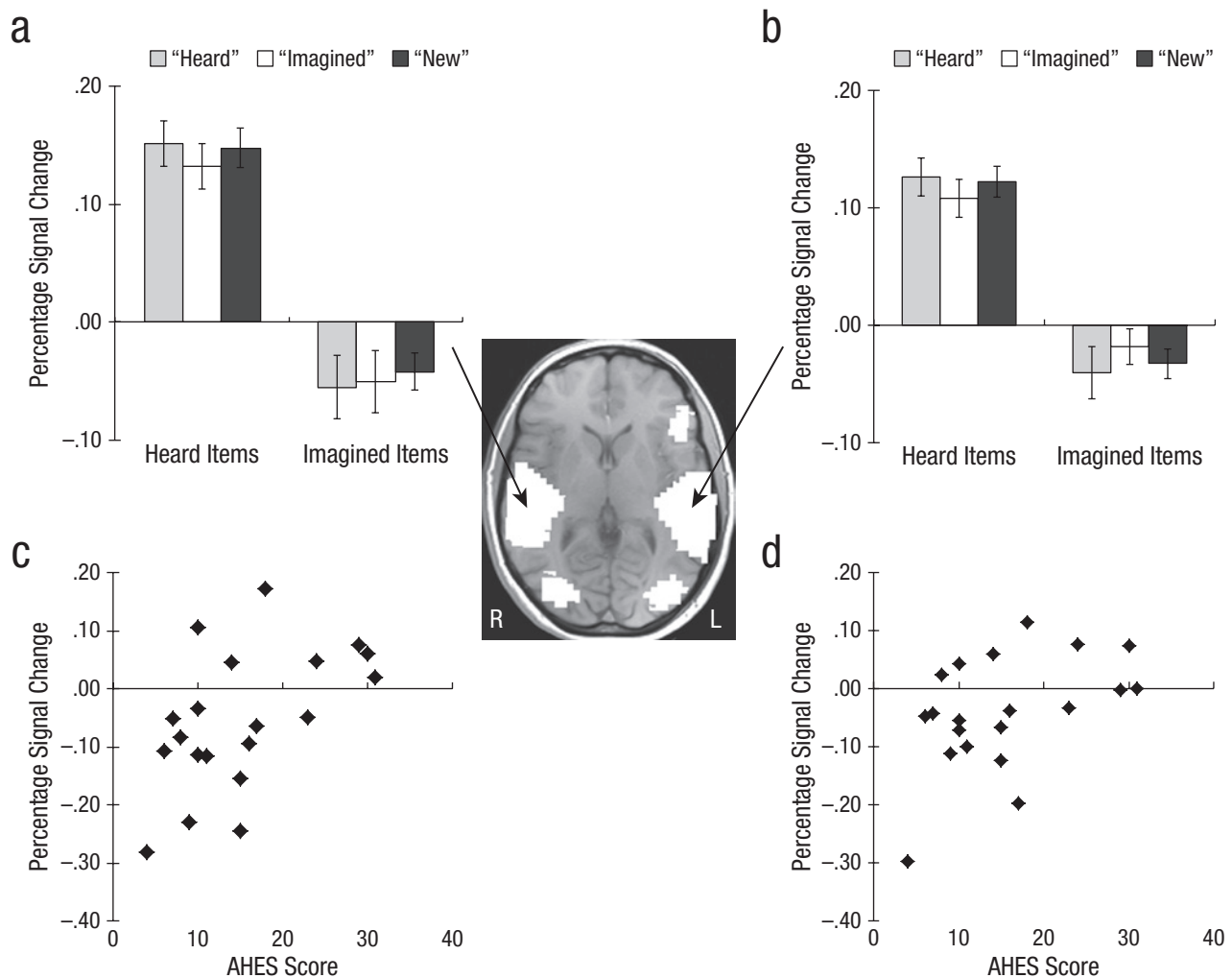

Fig. 3. Results for regions that activated when participants heard words and deactivated when they imagined hearing words in another person's voice. The brain image shows the areas of activation: bilateral superior temporal gyrus (STG) extending into middle and transverse temporal gyri and inferior parietal lobule (IPL; Brodmann's areas 22, 21, 39, 37). The graphs (a, b) show mean percentage change in blood-oxygen-level-dependent (BOLD) signal at encoding for heard and imagined items as a function of participants' response at test. Error bars show standard errors of the mean. The scatter plots (c, d) show mean percentage change in BOLD signal at encoding for imagined items called "heard" as a function of score on the Auditory Hallucination Experience Scale (AHES). Results for the right hemisphere are shown in the left column, and results for the left hemisphere are shown in the right column. R = right, L = left.

frequently seen in working memory tasks, especially those that require manipulation of information (e.g., Hanakawa et al., 2002; see Cabeza & Nyberg, 2000, for a review), which suggests that this area is associated with engaging cognitive operations. Hence, the pattern is consistent with the prediction from the SMF (Johnson et al., 1993) that records of cognitive operations carried out at the time of encoding (in this case, imagining) can later be used to judge an event as self-generated rather than perceived.[3]

Another (perhaps more speculative) possibility consistent with the SMF is that, in this paradigm, activity in this region of MFG represents cognitive operations specifically involved in localizing active auditory representations or the representation of location information. In effect, reality monitoring of speech requires "location" cues that differentiate internal from external sources. Further, it has been suggested that hallucinations reflect difficulty with the "subjective spatial location" of imagined events (Woodward & Menon, 2013, p. 171). There is evidence for a dorsal "where" auditory PFC pathway that is active during sound localization in both perceptual and working memory tasks (see Arnott & Alain, 2011, for a review). For example, an area with a local maximum about 1 voxel from our BA 6 region showed increased activity in an *n*-back task requiring participants to respond on the basis of sound location rather than sound identity (Leung & Alain, 2011). The current BA 6 is also similar to an area found during the delay period of a working memory task when participants had to hold auditory locations in mind versus a saccade localizer (Tark & Curtis, 2009) and when participants did a sound location delayed-match-to-sample task versus a sound identity-match-to-sample task (Arnott, Grady, Hevenor, Graham, & Alain, 2005).

The fact that left IFG (BA 45, 44) was more active at encoding for imagined than for heard words is consistent with the association of this area with speech generation (e.g., Horwitz et al., 2003); thus, one might expect the records of the cognitive operations associated with generating the auditory images to provide cues that items were imagined. However, encoding activity in this area was greater for items later called "heard" than "imagined" regardless of the source of the item; records of the activity in this region obviously were not serving as a cue that information was self-generated. As noted in the introduction, IFG may represent information generated (e.g., semantic information) as well as (or, in some subregions, instead of) the mode by which it is derived (see Mitchell et al., 2008). In fact, left IFG is associated not only with language generation but also with language comprehension (Skipper, Goldin-Meadow, Nusbaum, & Small, 2007) and semantic processing (see Gabrieli, Poldrack, & Desmond, 1998, for a review). Furthermore, encoding activity in a similar area was associated with subsequent

true *and* false memory in a study using a procedure designed to induce false memories for semantic associates of presented words (Kim & Cabeza, 2007). According to the SMF, greater semantic information is usually associated with perceived than with imagined events (Johnson & Raye, 1981). Thus, it is possible that source judgments at test were primarily sensitive to the amount of semantic information activated and not the amount of cognitive operations information. Clearly, the conditions determining the relative roles of IFG in cognitive operations as opposed to representation of semantic information during source monitoring remain to be specified as the functional and structural heterogeneity of IFG is clarified (cf. Fedorenko, Duncan, & Kanwisher, 2012).

In any event, the pattern of brain activity highlights the added value of neuroimaging data. Although, for imagined items, participants' behavioral "imagined" and "heard" responses did not differ, "heard" responses were related to IFG activity at encoding and "imagined" responses to MFG activity. These are areas one would expect to be associated with the processing of semantic and cognitive operations information, respectively. Hence, even with no difference in "heard" and "imagined" judgments, the brain data suggest that, as predicted by the SMF, participants were likely basing these reality-monitoring attributions on the relative value of different types of information.

Support for this proposition comes also from the AHES correlations with STG activity (an auditory area) only for imagined items called "heard." Bilateral STG (extending to adjacent areas) activated when participants heard words and, on average, it deactivated when they imagined hearing words in another person's voice. Assuming that STG represents auditory information, the SMF would predict that differences in activity for heard and imagined items should provide a basis for reality-monitoring judgments. Thus, increased activity in this region during imagination would decrease the availability of an important cue about the origin of auditory information. If so, activity in STG during auditory imagination may provide a useful marker of proneness to hallucinations. Consistent with this idea, one of our most interesting findings was that scores on the AHES were positively related to activity in these areas for imagined items that participants subsequently incorrectly called "heard." It may be that people prone to auditory hallucinations are good auditory imagers—that is, they relatively effortlessly or spontaneously produce vivid auditory imaginations that rival those of actually heard words. In clinical populations, individuals with schizophrenia who experience auditory verbal hallucinations are more likely to confuse what they have heard and what they imagined hearing in a laboratory reality-monitoring task than those without a history of auditory verbal hallucinations (Brunelin et al., 2006). Moreover, STG activity is associated with auditory hallucinations in patients (e.g.,

Bentaleb, Beauregard, Liddle, & Stip, 2002; Dierks et al., 1999; van de Ven et al., 2005), and hyperactivity of STG, perhaps due to reduced control from PFC, has been at the center of several theories of auditory hallucinations (see Woodward & Menon, 2013, for a recent review and discussion). Consistent with predictions based on the SMF that differences in activity in perceptual-processing areas should provide a basis for reality-monitoring judgments, a recent meta-analysis (Kompus et al., 2011) showed that patients with auditory hallucinations demonstrate both decreased activation in STG during external (i.e., perceptual) auditory stimulation and increased activation in the absence of external auditory stimulation.

In summary, the current findings provide new evidence regarding the processes affecting auditory reality monitoring and source misattribution of auditory information. We found that brain activity during encoding of words predicted whether people later mistakenly responded "heard" when presented with imagined items. Consistent with the SMF, results showed that less activity in a brain area associated with cognitive operations (left MFG) and more activity in a brain area associated with semantic processing (left IFG) during imagining another person's voice led healthy people to later judge imagined items as "heard." Further, the more participants were prone to have auditory-hallucination-like experiences, the more brain areas associated with speech perception (right and left STG, BA 22) were active for imagined items subsequently called "heard." The findings thus contribute to the understanding of the neural correlates of reality-monitoring errors for auditory information.

Author Contributions

E. Sugimori developed the study concept. E. Sugimori, M. K. Johnson, K. J. Mitchell, and C. L. Raye contributed to the study design. E. Sugimori and E. J. Greene prepared the materials, and E. Sugimori programmed the experiment, with input from M. K. Johnson, K. J. Mitchell, and C. L. Raye. Data were collected by E. Sugimori and K. J. Mitchell. E. J. Greene performed the functional MRI analyses, and E. Sugimori performed the behavioral data analyses under the supervision of M. K. Johnson and K. J. Mitchell. E. Sugimori, M. K. Johnson, K. J. Mitchell, and C. L. Raye contributed to the interpretation and the writing of the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgments

We thank the magnetic-resonance technicians at the Yale Magnetic Resonance Research Center for help in data collection, and Matt Sazma for help with figures.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. The shape condition was included in the initial ANOVA so that areas related to both heard and imagined words might be identified; however, there were no areas where activations for heard and imagined words were equal. The shape condition was of no further interest, and those data were not included in the subsequent memory analyses.
2. There also was no difference between "heard" and "imagined" responses for new items, which argues against an overall bias to call items "heard" whenever participants were not sure about the source.
3. We did not find differences in the medial PFC area associated with recollection of cognitive operations in other reality-monitoring studies (e.g., Turner, Simons, Gilbert, Frith, & Burgess, 2008; Vinogradov et al., 2006). One possible reason may be related to the auditory imagination requirements of our task versus other kinds of generation (e.g., solving anagrams) used in other studies. Another possibility is that medial PFC may come on-line during monitoring at test, not during initial generation of cognitive operations. Studies aimed at reconciling these differences should be informative.

References

Arnott, S. R., & Alain, C. (2011). The auditory dorsal pathway: Orienting vision. *Neuroscience and Biobehavioral Reviews*, *35*, 2162–2173.

Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., & Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. *Journal of Cognitive Neuroscience*, *17*, 819–831.

Bentaleb, L. A., Beauregard, M., Liddle, P., & Stip, E. (2002). Cerebral activity associated with auditory verbal hallucinations: A functional magnetic resonance imaging case study. *Journal of Psychiatry and Neuroscience*, *27*, 110–115.

Brunelin, J., Combris, M., Poulet, E., Kallel, L., D'Amato, T., Dalery, J., & Saoud, M. (2006). Source monitoring deficits in hallucinating compared to non-hallucinating patients with schizophrenia. *European Psychiatry*, *21*, 259–261.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*, 1–47.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology A*, *33*, 497–505. doi:10.1080/14640748108400805

Dierks, T., Linden, D. E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., & Singer, W. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Neuron*, *22*, 615–621.

Dobbins, I. G., & Wagner, A. D. (2005). Domain-general and domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cerebral Cortex*, *15*, 1768–1778.

Dobson, M., & Markham, R. (1993). Imagery ability and source monitoring: Implications for eyewitness memory. *British Journal of Psychology*, *84*, 111–118.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*, 2059–2062.

Finke, R. A., Johnson, M. K., & Shyi, G. C. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, *16*, 133–137.

Foley, M. A., Johnson, M. K., & Raye, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development*, *54*, 51–60.

Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences, USA*, *95*, 906–913.

Gonsalves, B., Reber, P. J., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Paller, K. A. (2004). Neural evidence that vivid imagining can lead to false remembering. *Psychological Science*, *15*, 655–660.

Hanakawa, T., Honda, M., Sawamoto, N., Okada, T., Yonekura, Y., Fukuyama, H., & Shibasaki, H. (2002). The role of rostral Brodmann area 6 in mental-operation tasks: An integrative neuroimaging approach. *Cerebral Cortex*, *12*, 1157–1170.

Horwitz, B., Amunts, K., Bhattacharyya, R., Patkin, D., Jeffries, K., Zilles, K., & Braun, A. R. (2003). Activation of Broca's area during the production of spoken and signed language: A combined cytoarchitectonic mapping and PET analysis. *Neuropsychologia*, *41*, 1868–1876.

Johnson, M. K. (1988). Discriminating the origin of information. In T. F. Oltmanns & B. A. Maher (Eds.), *Delusional beliefs* (pp. 34–65). New York, NY: Wiley & Sons.

Johnson, M. K. (2006). Memory and reality. *American Psychologist*, *61*, 760–771.

Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, *16*, 337–342.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.

Johnson, M. K., Raye, C. L., Mitchell, K. J., & Ankudowich, E. (2011). The cognitive neuroscience of true and false memories. In R. F. Belli (Ed.), *Nebraska Symposium on Motivation: Vol. 58. True and false recovered memories: Toward a reconciliation of the debate* (pp. 15–52). New York, NY: Springer.

Johnson, M. K., Raye, C. L., Wang, A. Y., & Taylor, T. H. (1979). Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 229–240.

Kensinger, E. A., & Schacter, D. L. (2005). Emotional content and reality-monitoring ability: fMRI evidence for the influences of encoding processes. *Neuropsychologia*, *43*, 1429–1443.

Kensinger, E. A., & Schacter, D. L. (2006). Neural processes underlying memory attribution on a reality-monitoring task. *Cerebral Cortex*, *16*, 1126–1133.

Kim, H., & Cabeza, R. (2007). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cerebral Cortex*, *17*, 2143–2150.

Kompus, K., Westerhausen, R., & Hugdahl, K. (2011). The "paradoxical" engagement of the primary auditory cortex in patients with auditory verbal hallucinations: A meta-analysis of functional neuroimaging studies. *Neuropsychologia*, *49*, 3361–3369.

Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., Toga, A. W., & Mazziotta, J. C. (1997). Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping*, *5*, 238–242.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A., & Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*, 120–131.

Larøi, F., Van der Linden, M., & Marczewski, P. (2004). The effects of emotional salience, cognitive effort and meta-cognitive beliefs on a reality monitoring task in hallucination-prone subjects. *British Journal of Clinical Psychology*, *43*, 221–233.

Launay, G., & Slade, P. (1981). The measurement of hallucinatory predisposition in male and female prisoners. *Personality and Individual Differences*, *2*, 221–234.

Leung, A. W. S., & Alain, C. (2011). Working memory load modulates the auditory "what" and "where" neural networks. *NeuroImage*, *55*, 1260–1269.

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*, 423–428.

Lindsay, D. S. (2008). Source monitoring. J. H. Byrne (Series Ed.) & H. L. Roediger, III (Vol. Ed.), *Learning and memory: A comprehensive reference* (Vol. 2, pp. 325–347). Oxford, England: Elsevier.

Mitchell, K. J., Raye, C. L., McGuire, J. T., Frankel, H., Greene, E. J., & Johnson, M. K. (2008). Neuroimaging evidence for agenda-dependent monitoring of different features during short-term source memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 780–790.

Simons, J. S., Gilbert, S. J., Owen, A. M., Fletcher, P. C., & Burgess, P. W. (2005). Distinct roles for lateral and medial anterior prefrontal cortex in contextual recollection. *Journal of Neurophysiology*, *94*, 813–820.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, *101*, 260–277.

Sugimori, E., Asai, T., & Tanno, Y. (2011). Sense of agency over speech and proneness to auditory hallucinations: The reality monitoring paradigm. *The Quarterly Journal of Experimental Psychology*, *64*, 169–185.

Tark, K.-J., & Curtis, C. E. (2009). Persistent neural activity in the human frontal cortex when maintaining space that is off the map. *Nature Neuroscience*, *12*, 1463–1468.

Turner, M. S., Simons, J. S., Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2008). Distinct roles for lateral and medial rostral prefrontal cortex in source monitoring of perceived and imagined events. *Neuropsychologia*, *46*, 1442–1453.

van de Ven, V. G., Formisano, E., Röder, C. H., Prvulovic, D., Bittner, R. A., Dietz, M. G., . . . Linden, D. E. (2005). The spatiotemporal pattern of auditory cortical responses during verbal hallucinations. *NeuroImage*, *27*, 644–655.

Vinogradov, S., Luks, T. L., Simpson, G. V., Schulman, B. J., Glenn, S., & Wong, A. E. (2006). Brain activation patterns during memory of cognitive agency. *NeuroImage*, *31*, 896–905.

Woods, R. P., Cherry, S. R., & Mazziotta, J. C. (1992). Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, *16*, 620–633.

Woodward, T. S., & Menon, M. (2013). Misattribution models: II. Source monitoring in hallucinating schizophrenia subjects. In R. Jardri, A. Cauchia, P. Thomas, & D. Pins (Eds.), *The neuroscience of hallucinations* (pp. 169–184). New York, NY: Springer.