

Finding the imposter: brain connectivity of lesions causing delusional misidentifications

R. Ryan Darby,^{1,2,3,*} Simon Laganier,^{1,*} Alvaro Pascual-Leone,¹ Sashank Prasad⁴ and Michael D. Fox^{1,2,5}

*These authors contributed equally to this work.

See McKay and Furl (doi:10.1093/aww323) for a scientific commentary on this article.

Focal brain injury can sometimes lead to bizarre symptoms, such as the delusion that a family member has been replaced by an imposter (Capgras syndrome). How a single brain lesion could cause such a complex disorder is unclear, leading many to speculate that concurrent delirium, psychiatric disease, dementia, or a second lesion is required. Here we instead propose that Capgras and other delusional misidentification syndromes arise from single lesions at unique locations within the human brain connectome. This hypothesis is motivated by evidence that symptoms emerge from sites functionally connected to a lesion location, not just the lesion location itself. First, 17 cases of lesion-induced delusional misidentifications were identified and lesion locations were mapped to a common brain atlas. Second, lesion network mapping was used to identify brain regions functionally connected to the lesion locations. Third, regions involved in familiarity perception and belief evaluation, two processes thought to be abnormal in delusional misidentifications, were identified using meta-analyses of previous functional magnetic resonance imaging studies. We found that all 17 lesion locations were functionally connected to the left retrosplenial cortex, the region most activated in functional magnetic resonance imaging studies of familiarity. Similarly, 16 of 17 lesion locations were functionally connected to the right frontal cortex, the region most activated in functional magnetic resonance imaging studies of expectation violation, a component of belief evaluation. This connectivity pattern was highly specific for delusional misidentifications compared to four other lesion-induced neurological syndromes ($P < 0.0001$). Finally, 15 lesions causing other types of delusions were connected to expectation violation ($P < 0.0001$) but not familiarity regions, demonstrating specificity for delusion content. Our results provide potential neuroanatomical correlates for impaired familiarity perception and belief evaluation in patients with delusional misidentifications. More generally, we demonstrate a mechanism by which a single lesion can cause a complex neuropsychiatric syndrome based on that lesion's unique pattern of functional connectivity, without the need for pre-existing or hidden pathology.

- 1 Berenson-Allen Center for Non-Invasive Brain Stimulation and Division of Cognitive Neurology, Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
- 2 Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
- 3 Department of Neurology, McLean Psychiatric Hospital, Harvard Medical School, Belmont, MA, USA
- 4 Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
- 5 Athinoula A. Martinos Centre for Biomedical Imaging, Charlestown, MA, USA

Correspondence to: Ryan Darby,
Berenson-Allen Center for Noninvasive Brain Stimulation,
Beth Israel Deaconess Medical Center,
330 Brookline Ave, Kirstein Building KS 158,
Boston, MA 02215, USA
E-mail: rdarby@bidmc.harvard.edu

Keywords: Capgras; fregoli; delusions; delusional misidentifications; connectivity

Introduction

Delusional misidentification syndromes are among the most striking and least understood syndromes encountered in neurology and psychiatry (for an example of a patient's description, see Supplementary Video 1). In Capgras syndrome, a patient is able to identify and recognize a familiar person, such as a spouse or parent, but experiences that person as unfamiliar, leading to the bizarre conclusion that their family member has been replaced by an imposter (Capgras and Reboul-Lachaux, 1923). In Fregoli syndrome, a stranger is believed to be a familiar person in disguise (Courbon and Fail, 1927). Similar delusional misidentifications can occur for non-human animals (Darby and Caplan, 2016) and personally relevant locations and buildings, such as one's home (Pick, 1903).

Many theories have attempted to explain how the abnormal belief in Capgras syndrome in particular emerges. In his original report, Capgras proposed that dysfunction of brain regions involved in the experience of familiarity must be required (Capgras and Reboul-Lachaux, 1923). It remains unclear which brain regions are involved in this abnormal sense of familiarity, although dysfunction of the dorsal visual pathway (Ellis and Young, 1990) and disconnection between the fusiform face area and amygdala (Hirstein and Ramachandran, 1997; Ramachandran, 1998) have been proposed. In addition to impaired familiarity perception, other theories have proposed deficits in theory of mind (Hirstein, 2010), personal relatedness (Feinberg, 2011), or autobiographical memory (Staton *et al.*, 1982; Darby and Caplan, 2016).

In each of these explanations, it is assumed that there is an additional functional deficit that allows the abnormal perception to go unchallenged, leading to the development of a delusional belief. According to one model, all delusions are conceptualized as a 'two-hit' process, with (i) an abnormal perception leading to the specific bizarre content for a given delusion; and (ii) impaired belief evaluation, which allows the abnormal delusional belief to form (Coltheart, 2007, 2010). Other models propose that belief and perception are integrated into a single process, such as prediction error (Corlett *et al.*, 2010), or that belief evaluation and perception interact in a bidirectional manner (Young, 2008). While each theory proposes abnormal familiarity perception and belief evaluation, a neurobiological explanation for how this might occur in patients with delusional misidentifications is lacking.

Patients with focal brain lesions offer the opportunity to test these theories by linking damage to specific neuroanatomical locations with delusional misidentifications (Feinberg, 2005; Devinsky, 2009; Darby and Prasad, 2016). However, this approach has been difficult for three reasons. First, lesions causing delusional misidentifications are rare and have been reported across different brain areas (Darby and Prasad, 2016), making localization to a single region difficult. Second, the most common

location for these lesions is the right frontal lobe (Darby and Prasad, 2016), an area without a clear role in processing familiarity (Gobbini and Haxby, 2007). Third, and perhaps most challenging, it is difficult to explain how a single brain lesion could disrupt multiple different network functions as required by the 'two hit' and other theories (Coltheart, 2007, 2010). Due in part to these difficulties, it has been suggested that only patients with pre-existing psychiatric disease (Devine *et al.*, 2014), dementia (Levine and Grek, 1984; Rabins *et al.*, 1991), hidden brain lesions (Hirstein and Ramachandran, 1997), or other predisposing factors (Ellis and Young, 1990) are susceptible to lesion-induced delusional misidentifications.

Recently, we developed and validated a technique termed lesion network mapping for investigating lesion-induced neurological symptoms (Boes *et al.*, 2015). This technique enables one to test the hypothesis that symptoms emerge from sites functionally connected to a lesion location, not just the lesion location itself (Monakow, 1914). Here, we apply this technique towards understanding lesion-induced delusional misidentifications. We hypothesized that lesion locations causing delusional misidentifications would be connected to regions involved in familiarity processing and belief evaluation.

Materials and methods

Patient cases from the literature

To identify patients with delusional misidentifications for persons or places, we searched Pubmed for the terms related to delusional misidentification syndromes ('delusional misidentification syndrome', DMS, Capgras, Fregoli, 'reduplicative paramnesia', subjective doubles, intermetamorphosis, or Cotard) and evidence of abnormal neuroimaging (MRI or CT) or neurological injury (stroke, haemorrhage, trauma or lesion). Patients were required to have a delusional belief involving the sense of under-familiarity for a person (Capgras) or place (reduplicative amnesia), or the feeling of over-familiarity for a person (Fregoli) or place (reduplicative paramnesia). Inclusion criteria included clinical data sufficient to determine the nature of the delusion, an acute, localized neurological injury (ischaemic or haemorrhagic stroke), and imaging showing lesions with enough clarity to trace onto a standard brain atlas. Patients with pre-existing psychiatric or neurological disease were excluded. Fifteen patients were identified using these criteria (Supplementary Table 1).

Patient cases at the study centre

In addition, we included two cases of patients encountered in our centre.

Patient 1

An 83-year-old female initially presented with sudden onset of left-sided face, arm, and leg weakness and confusion. She was found to have a large right frontal haemorrhage, which was stable on serial imaging. Her weakness slowly improved,

although she was noted to have difficulty in orientation to place, believing she was in a different city while hospitalized, and at times believing she was at her place of work while at a rehabilitation hospital. Four months after her initial presentation she presented for follow-up to the neurology clinic where she was noted to have the persistent delusion that her home was not her ‘real’ home. She described recognizing individual objects in her home, its spatial layout, and other features as being identical with her actual home, but believed it was nevertheless a different home located in a different location. She would occasionally pack belongings, believing that she would be returning to her ‘real’ house. She denied other delusions, including misidentifications of persons or other objects. She had documented impairment in short-term memory at her initial presentation, but had normal cognitive and neurological exam at the time of her follow-up appointment. At follow-up an additional 4 months later, the delusion had resolved. See Supplementary Video 1 for patient interview.

Patient 2

A 70-year-old female presented initially with confusion. A more detailed history revealed that she had been at a friend’s house when she began saying that the house did not ‘feel’ like her friend’s real house, and must be a replica. Upon returning to her own home she felt that this home was not her real home, and must be in a different city despite looking identical. Her delusion resolved by the next day. She was found to have a right frontal ischaemic stroke on neuroimaging. She had mild left-sided face, arm, and leg weakness, disorientation to date, and short-term memory impairment on exam. Unfortunately, she had progressive right internal cerebral artery territory

infarctions from a progressive vasculopathy due to varicella-zoster virus, and died ~1 month after presentation. She had no further delusions documented during that time period.

Lesion localization

Published images or the patient’s own MRI were used to map each lesion location to a common brain atlas (Fig. 1). Lesions were traced by hand onto a standardized brain atlas ($2 \times 2 \times 2$ MNI space) using FSL. All available images were traced for each patient. In the case of our two personal, unpublished cases, each axial slice involving the lesion was traced to approximate the 3D volume of the lesion.

Lesion network mapping

Our group recently developed a technique termed lesion network mapping that identifies brain regions functionally connected to lesion locations causing a given neuropsychiatric symptom (Boes *et al.*, 2015). This technique avoids the need to perform functional brain imaging on the patients themselves and has been validated across four different neurological syndromes. Briefly, traced lesions were used as individual seeds in a resting state connectivity analysis with data from 98 normal subjects. The correlated time course between each lesion location and every other brain voxel was calculated using the resting state data from each individual normal control. These correlations for all 98 subjects were then combined to calculate a T-score value for every individual voxel. Both positive and negative correlations with the time course of the lesion location were included. Voxels were thresholded at $T > \pm 4.25$

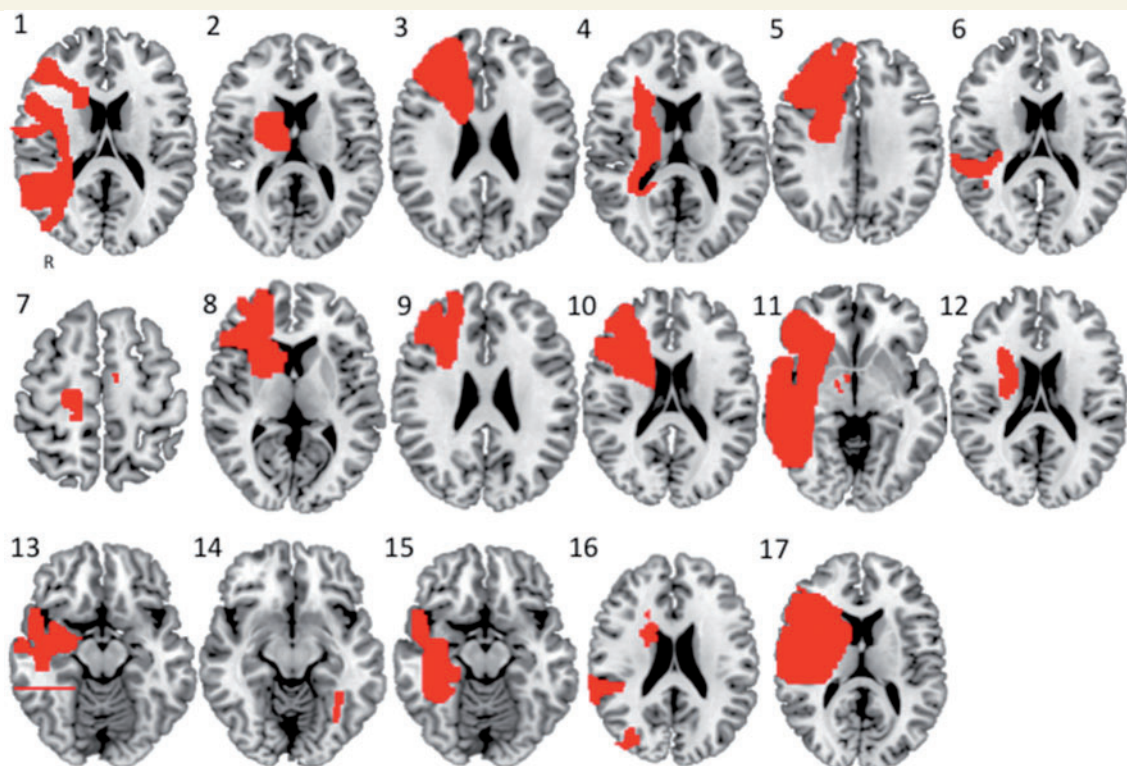


Figure 1 Lesions causing delusional misidentifications. Each lesion, numbered 1 through 17, was identified from a literature search or from cases seen by the authors and manually traced onto a common brain atlas (MNI).

($P < 0.00005$ for 98 degrees of freedom, uncorrected for multiple comparisons) in order to create a binarized map of significantly functionally connected regions to each patient's lesion site. Finally, maps from each of the patients were combined to form the lesion network mapping overlap for the group, showing the number of patients with lesions functionally connected with each individual voxel (Fig. 2).

Familiarity ALE meta-analysis

We searched Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed/>) for the terms 'personal or personally', familiar, and functional MRI, including studies where personally familiar stimuli (faces, names, voices, places, objects) were compared with unfamiliar stimuli. We included stimuli of different categories (persons, places, objects) and sensory modalities (visual, auditory) because delusional misidentifications have been reported in each of these domains. The coordinates of all significant foci from each study were included. If coordinates were reported in Talarach space, they were first converted to MNI space using the algorithm available on GingerALE.

The activation likelihood estimate of all foci was calculated using GingerALE v2.3.3 (www.brainmap.org). First, a 3D Gaussian probability distribution is created centred on each individual foci and modified by the sample size from each study in order to estimate the spatial uncertainty surrounding each focus. These distributions are then combined across all experiments to arrive at the activation likelihood estimate

maps. Finally, a cluster-level inference is used to determine significance. The true convergence on the ALE is compared against a null distribution of 1000 simulated datasets with identical number of foci, experiments, and subjects, but with the foci randomly distributed. Cluster-forming threshold was set at $P < 0.001$ and cluster-level inference threshold was set at $P < 0.05$.

Belief evaluation ALE meta-analyses

The process of detecting and challenging delusional thought content is difficult to define and test in normal subjects experimentally. Here, we define belief evaluation as the process of detecting and evaluating events that violate one's expectation, sometimes also referred to as prediction error (Corlett *et al.*, 2010). While belief evaluation is sometimes included as a type of 'reality monitoring', reality monitoring has also been used to refer to source monitoring of memory retrieval related to internally versus externally generated stimuli (Mitchell and Johnson, 2009; Metzack *et al.*, 2015). Therefore, we use the term 'belief evaluation' to avoid ambiguity with this second meaning of reality monitoring.

The most straightforward and widely studied task for expectation violation is the detection of invalid cues, a standard component of the Posner invalid cue paradigm (Posner, 1980). Though certainly not identical to belief evaluation, detecting events that violate expectation may be part of detecting and challenging delusions (Corlett *et al.*, 2010). We searched

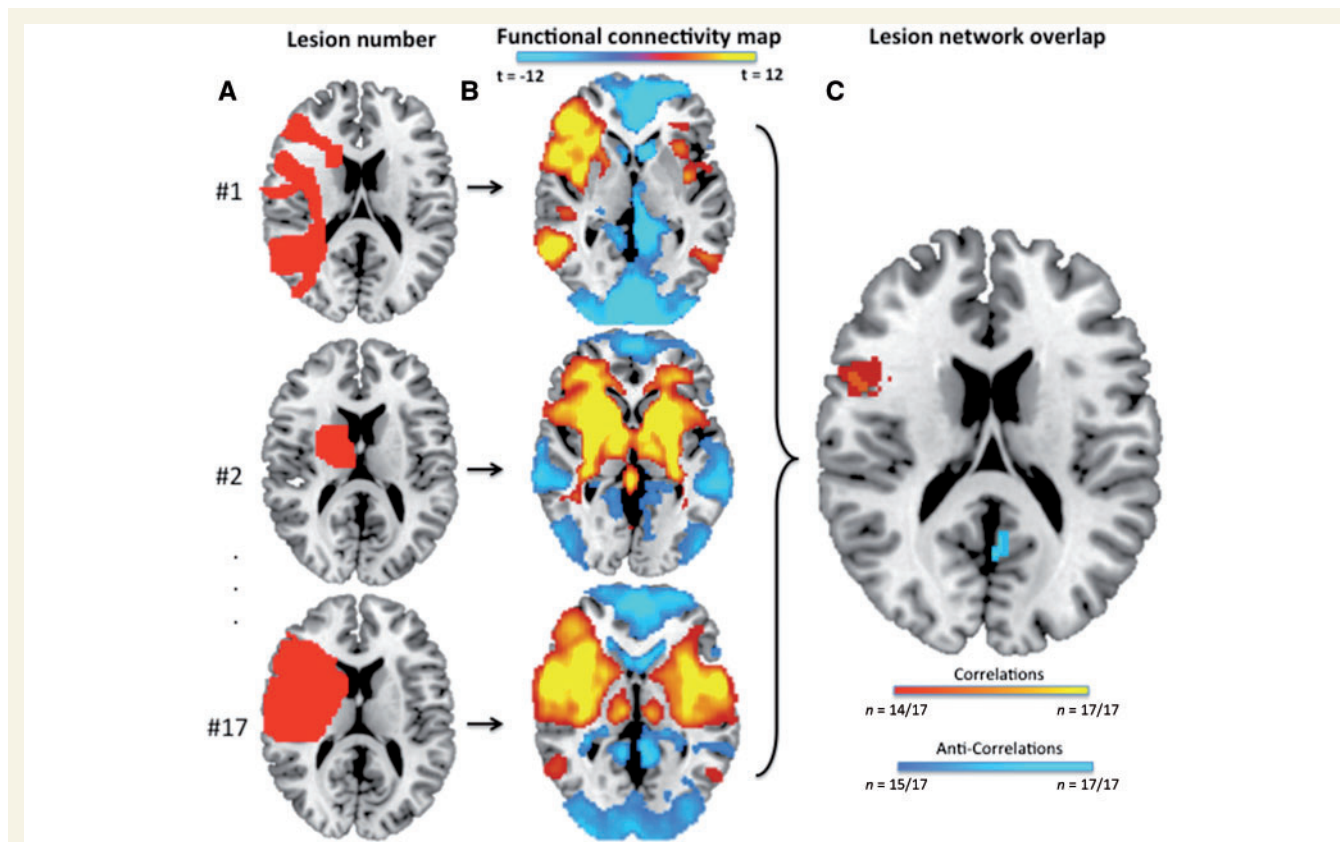


Figure 2 Lesion network mapping technique. (A) Lesions traced onto a standardized MNI brain template. (B) Brain regions functionally connected to each lesion location based on a large resting state functional connectivity database. (C) Overlap in the functional connectivity maps from each lesion identifies brain regions functionally connected to the greatest number of lesion locations.

Pubmed for the terms (invalidity or ‘violation of expectation’) and functional MRI, as well as relevant references from review articles. We included studies that used variations of the Posner invalid cue paradigm. This paradigm involves identifying the spatial location of target stimuli after explicitly being instructed that the stimuli will occur in a particular location. Inclusion in our meta-analysis required reporting significant coordinates from regions contrasting unexpected trials (where the stimuli occurred in the opposite location) with expected trials, with selection of the appropriate stimuli using a volitional motor response.

To ensure that results were not dependent on the details of this meta-analysis, we confirmed our findings using two other meta-analyses of expectation violation. First, we obtained the results of an independent, recently published ALE meta-analysis of auditory and visual ‘oddball’ tasks (Kim, 2014). These tasks also include violations of expectation, where an unexpected image or tone is placed within a sequence of expected images or tones.

Second, we performed a meta-analysis of two types of functional imaging studies involving expectation violation of cognitive beliefs. First, we included studies where expectations of a previously learned association are violated, leading to prediction error (Corlett *et al.*, 2004). For example, a subject learns that a certain food is associated with an allergic reaction. After learning this association, on some trials the food will not be associated with an allergic reaction, violating the expectation based on this previously learned belief (Corlett *et al.*, 2004). Second, we included studies of logical syllogisms where a logical conclusion violates the expectation based on a previously held belief. For example, the argument: ‘No addictive things are inexpensive. Some cigarettes are inexpensive. Therefore some cigarettes are not addictive’ (Goel and Dolan, 2003) is logically valid, but the conclusion violates a subject’s prior beliefs that cigarettes are addictive. In contrast to the initial meta-analyses, which assessed lower-level expectation violation (e.g. an invalid cue or odd-ball), this additional meta-analysis included studies of prediction error and logical syllogism tasks that assess for more complex violations of expectation related to beliefs.

Connectivity between lesions and meta-analysis regions

Next we determined whether lesion locations causing delusional misidentifications showed greater connectivity to regions involved in familiarity perception or expectation violation (as identified in our ALE meta-analyses) than lesion locations causing other neurological syndromes. For this analysis, functional MRI time courses were extracted from each lesion location and each ALE-derived region of interest. The Pearson’s correlation coefficient between time courses was computed for each subject in our normative 98-subject dataset. Resulting *r*-values were converted to a normal distribution using Fischer’s *r* to *z* transform and statistically compared using a two-tailed *t*-test. Lesion locations causing delusional misidentifications were compared to lesion locations causing four other neurological syndromes (*n* = 77) from our previously published work (Boes *et al.*, 2015). These syndromes included auditory hallucinations, visual hallucinations, post-stroke pain, and

subcortical aphasia. All statistics were computed using the statistical package STATA (College Station, TX, version 14.0).

Lesions causing other delusions

As a final, more stringent test of specificity, lesion locations causing delusional misidentifications were compared to lesion locations causing delusions with different content (e.g. persecution). We identified an additional 15 lesions from the literature that resulted in delusions other than delusional misidentifications (Kumral and Oztürk, 2004). The above lesion network mapping and time course analyses were repeated using this separate cohort of delusion lesions and results were statistically compared to lesions causing delusional misidentifications and the four other neurological syndromes.

Results

Lesion identification and network mapping

We identified 17 lesions resulting in delusional misidentification (Fig. 1).

Lesion network mapping was used to identify brain regions functionally connected to the greatest number of lesion locations (Fig. 2). The left retrosplenial cortex was the only site in the brain functionally connected with all 17 lesion locations (negative correlation, Fig. 3A). Sixteen of 17 lesions were also functionally connected (positively correlated) with a region in the right ventral frontal cortex/ anterior insula (Fig. 4A). Other sites of shared connectivity were apparent at a slightly lower threshold (Supplementary Table 2).

Familiarity meta-analysis

We identified 15 neuroimaging studies that compared personally familiar stimuli to unfamiliar stimuli in normal subjects (Supplementary Table 3). The left retrosplenial cortex was the region most activated by personally familiar stimuli (Fig. 3B), and all 17 lesions causing delusional misidentifications were functionally connected to this location (Fig. 3C). Other regions from our lesion network mapping also matched activation sites from the familiarity meta-analysis (Supplementary Table 4).

Violation of expectation meta-analysis

We found 11 neuroimaging studies that compared the detection of unexpected versus expected stimuli using variations of the Posner paradigm (Supplementary Table 5). The right ventral frontal cortex was the region most activated by trials involving unexpected or invalid cues (Fig. 4B and Supplementary Table 6), and 14 of 17 lesions causing delusional misidentifications were significantly functionally connected to this region (Fig. 4C). Right frontal regions

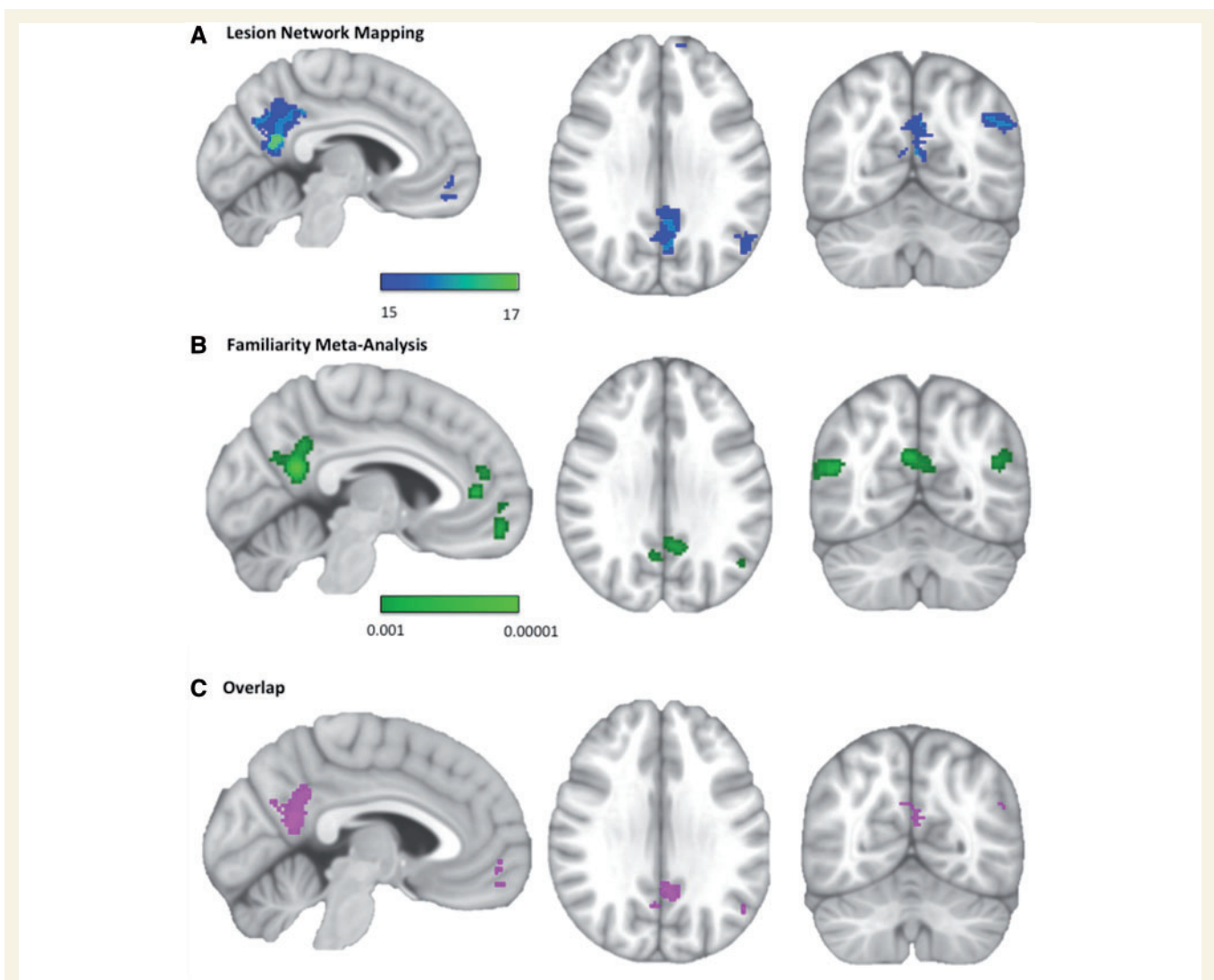


Figure 3 Network mapping of delusional misidentification lesions overlaps with regions involved in familiarity detection. (A) Regions functionally connected to lesion locations causing delusional misidentifications (negative correlations). Colour scale reflects the number of lesion locations with significant connectivity to each voxel. Peak coordinate: $x = -6$, $y = -56$, $z = 12$. **(B)** Regions most activated by familiar versus unfamiliar stimuli. Colour scale reflects the probability that a voxel is activated in neuroimaging studies of familiarity. Peak coordinate: $x = -4$, $y = -56$, $z = 20$. **(C)** Overlap image showing regions within the familiarity meta-analysis significantly connected to at least 15 of 17 lesions. Displayed brain slices from left to right are $x = -4$, $z = 30$, $y = -62$.

were also activated by expectation violation in ‘oddball’ tasks (Supplementary Fig. 1A; Kim, 2014) and expectation violation of cognitive beliefs (Supplementary Fig. 1B and Supplementary Tables 7 and 8).

Connectivity between lesions and meta-analysis regions

Although there was overlap between our lesion network and meta-analysis maps, a more direct test of our hypothesis is to perform a region of interest functional connectivity analysis. Specifically, lesion locations causing delusional misidentifications (Fig. 1) and voxels identified in our meta-

analyses (e.g. Figs 3B and 4B) were used as regions of interest and functional connectivity between regions was computed. Lesion locations causing delusional misidentifications were significantly negatively correlated with regions involved in familiarity detection (mean correlation = -0.23 , $P < 0.0001$, Fig. 5A) and significantly positively correlated with regions activated by violation of expectation (mean correlation = 0.24 , $P < 0.0005$, Fig. 5B). Results were similar using regions activated by expectation violation in the oddball task (mean correlation = 0.34 , $P < 0.0001$, Supplementary Fig. 2B) or expectation violation of beliefs (mean correlation = 0.16 , $P < 0.005$, Supplementary Fig. 2C). For all analyses, results were specific to lesions causing

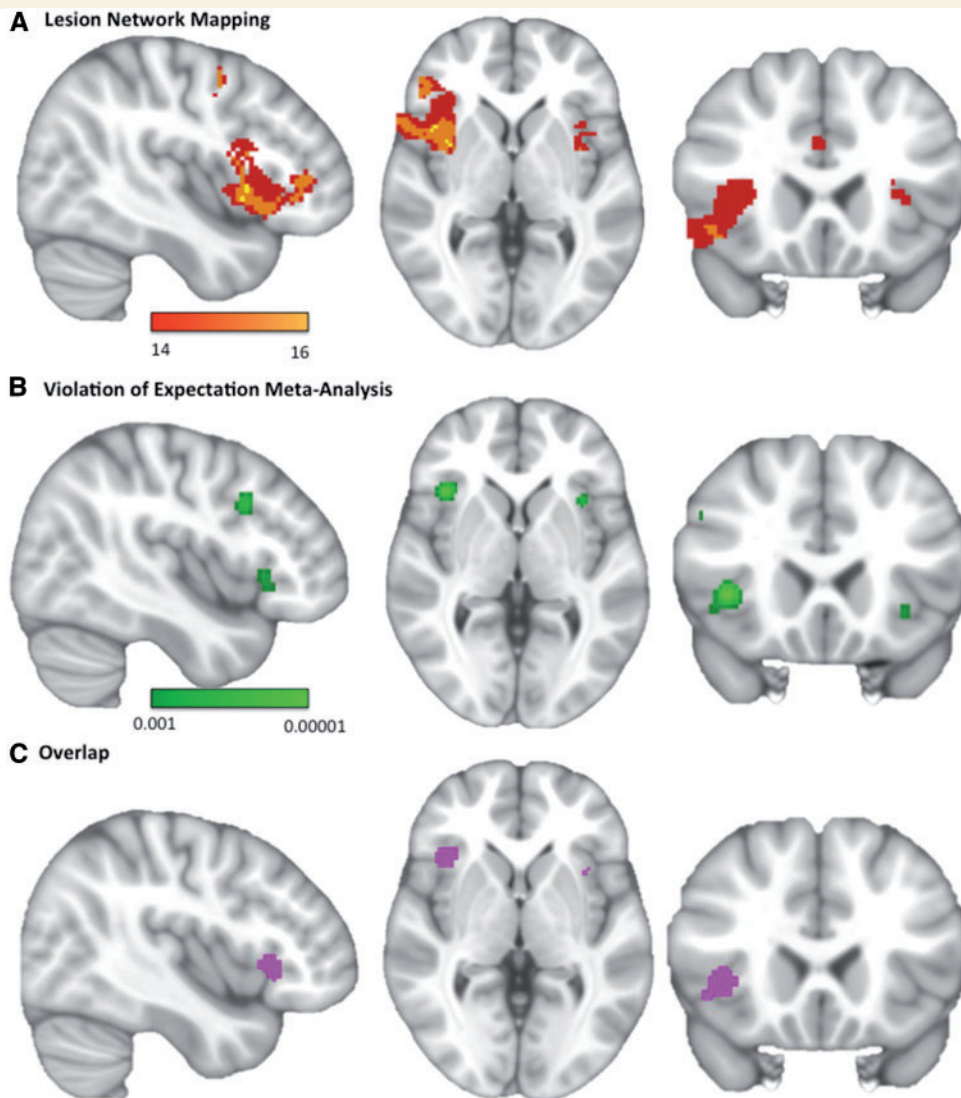


Figure 4 Network mapping of delusional misidentification lesions overlaps with regions involved in expectation violation. **(A)** Regions functionally connected to lesion locations causing delusional misidentifications (positive correlations). Colour scale reflects the number of lesion locations with significant connectivity to each voxel. Peak coordinate: $x = 54, y = 14, z = -10$. **(B)** Regions most activated by detection of invalid or unexpected stimuli in the Posner Paradigm. Colour scale reflects the probability that a voxel is activated in neuroimaging studies of invalid cues. Peak coordinate: $x = 38, y = 20, z = 2$. **(C)** Overlap image showing regions within the violation of expectation meta-analysis significantly connected to at least 14 of 17 lesions. Displayed brain slices from left to right are $x = 42, z = 22, y = 0$.

delusional misidentification: functional connectivity to meta-analysis regions was significantly stronger for lesions causing delusional misidentifications compared to lesions causing other neurological syndromes ($P < 0.0001$ for all comparisons).

Comparison to lesions causing other delusions

We next tested the hypothesis that connectivity of lesions to expectation violation areas is related to delusion formation more generally, while connectivity to familiarity perception areas is specific to delusional misidentifications. We identified an additional 15 lesions from the literature that

resulted in delusions other than delusional misidentifications (Kumral and Oztürk, 2004). Lesion network mapping identified overlap in the right frontal cortex similar to our initial cohort, but no overlap in the retrosplenial cortex (Supplementary Fig. 3 and Supplementary Table 9). Delusion lesion locations were not connected to familiarity regions (mean correlation = $-0.007, P = 0.89$, Fig. 6A) but were connected to regions involved in expectation violation using invalid cues (mean correlation = $0.13, P < 0.01$, Fig. 6B), oddball tasks (mean correlation = $0.18, P < 0.05$, Supplementary Fig. 4B), and belief tasks (mean correlation = $0.12, P < 0.005$, Supplementary Fig. 4C). Compared to control lesions that do not cause delusions, delusion lesions had stronger connectivity to regions

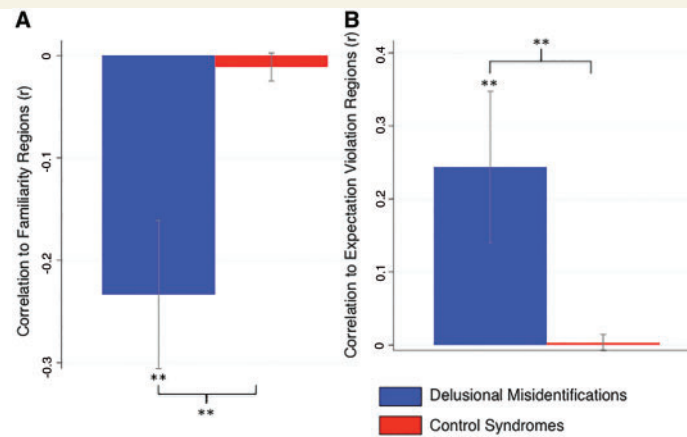


Figure 5 Lesions causing delusional misidentifications are functionally connected to brain regions involved in familiarity detection and violation of expectation. The temporal correlation in spontaneous functional MRI activity between lesion locations and regions identified in our familiarity meta-analyses (A) or violation of expectation meta-analysis (B) were computed using a cohort of healthy subjects. Correlations were averaged across our 17 lesion locations causing delusional misidentifications (blue) and compared to 77 lesions causing other neurological syndromes (control syndromes, red). * $P < 0.0005$. ** $P < 0.0001$.

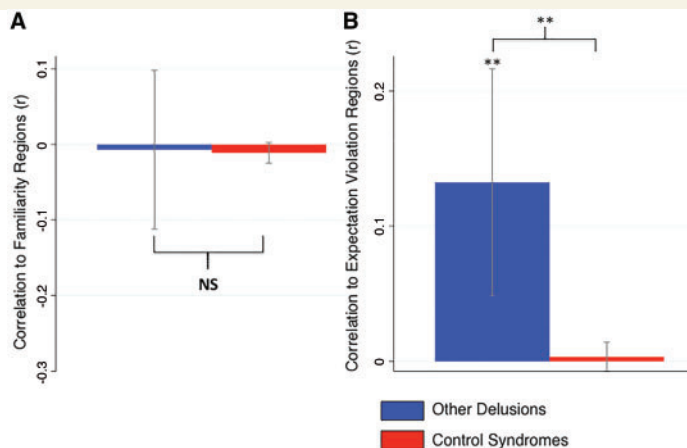


Figure 6 Lesions causing other delusions are functionally connected to brain regions involved in violation of expectation but not familiarity detection. The temporal correlation in spontaneous functional MRI activity between lesion locations and regions identified in our familiarity meta-analyses (A) or violation of expectation meta-analysis (B) were computed using a cohort of healthy subjects. Correlations were averaged across our 15 lesion locations causing other delusions (blue) and compared to lesions causing other neurological syndromes (red). * $P < 0.01$. ** $P < 0.0001$.

involved in expectation violation using invalid cues ($P < 0.0001$), oddball tasks ($P < 0.005$), and belief tasks ($P < 0.0001$), but connectivity to familiarity regions showed no difference ($P = 0.88$, Fig. 6A).

Direct comparison between lesions causing delusional misidentifications and lesions causing other delusions showed that delusional misidentifications were significantly more connected (negatively correlated) to regions involved in familiarity ($P = 0.001$). In contrast, there was no significant difference in connectivity to regions involved in expectation violation based on invalid cues ($P = 0.11$, not

significant), oddball tasks ($P = 0.09$, not significant), or belief tasks ($P = 0.46$, not significant).

Discussion

Our results demonstrate that lesions causing delusional misidentifications are characterized by a unique pattern of functional brain connectivity to both regions involved in perceiving familiarity and regions involved in expectation violation. This pattern of connectivity was specific

compared with other neurological syndromes. Lesions causing other types of delusions were not connected to familiarity regions (explaining the specificity of delusion content), but were connected to expectation violation regions. Taken together, our results provide a neuro-anatomical framework for understanding delusional misidentifications such as Capgras syndrome.

Neuroanatomical localization of familiarity

The retrosplenial cortex was functionally connected to all 17 lesion locations causing delusional misidentifications, was not connected to lesion locations causing other delusions (or other neurological syndromes), and was the brain area most associated with familiarity perception in our imaging meta-analysis. Prior theories have attributed the abnormal belief content in delusional misidentifications to deficits in determining the personal relatedness of objects (Feinberg, 2011), theory of mind (Hirstein, 2010), and retrieval of relevant autobiographical memories (Staton *et al.*, 1982; Darby and Caplan, 2016). All these functions fit proposed functions of the retrosplenial cortex (Vann *et al.*, 2009). Our results fail to support the specific neuro-anatomical localization proposed by prior models (Ellis and Young, 1990; Hirstein and Ramachandran, 1997; Ramachandran, 1998), but do support these model's hypothesis implicating regions involved in processing familiarity rather than those involved in facial and object recognition.

Ideally, our study would have included a group of control lesions that disrupt familiarity perception without causing delusions, similar to our group of control lesions causing delusions without disrupting familiarity. However, it's unclear that such a control group exists. Patients with ventromedial prefrontal cortex lesions have abnormal galvanic skin responses (GSR) to familiar faces (Tranel *et al.*, 1995; Tranel, 2000) similar to patients with Capgras syndrome (Ellis *et al.*, 1997; Hirstein and Ramachandran, 1997; Brighetti *et al.*, 2007), and do not have delusions, making them a potential control group (Coltheart, 2007, 2010). However, individual lesion locations for these patients were not provided in previous publications. More importantly, these patients do not lose their experience of familiarity (Tranel *et al.*, 1995; Young, 2009) and have abnormal GSR in tasks unrelated to familiarity (Bechara *et al.*, 1994, 1997; Moretto *et al.*, 2010), making it unlikely that they share the same perceptual deficit as patients with delusional misidentifications.

Neuroanatomical localization of belief evaluation

The right frontal cortex was functionally connected to almost all lesion locations causing delusions regardless of the specific delusion content, was not connected to lesion

locations causing other neurological syndromes, and was activated in all three of our expectation violation meta-analyses. The right ventral frontal cortex has been implicated in reorienting attention (Corbetta *et al.*, 2008) and contextual updating of one's internal representation of the external world (Geng and Vossel, 2013). The right dorsal frontal cortex has been implicated in belief evaluation and prediction error (Corlett *et al.*, 2004). Both regions are more connected to lesions causing delusions than lesions causing other neurological syndromes in our study. As such, these regions may play a role in lesion-induced delusion formation consistent with prior hypotheses (Coltheart, 2010; Corlett *et al.*, 2010).

Implications for theoretic models of delusional misidentification syndromes

Our results support the two-factor model of delusions (Coltheart, 2007, 2010) by showing that (i) lesions causing delusional misidentifications are functionally connected to both familiarity and belief evaluation regions; and (ii) connectivity to familiarity regions was specific to delusional misidentifications and not other delusions. Our results also support the prediction error model (Corlett *et al.*, 2010; Corlett and Fletcher, 2015) by showing that lesions are connected to right frontal regions involved in expectation violation. Finally, our results could be consistent with the interactionist model (Young, 2008) as lesion connectivity to both familiarity regions and expectation violation regions may disrupt integration between these processes.

Our results may also contradict aspects of prior theories. For example, it is unclear how connectivity to familiarity regions in the retrosplenial cortex fits with prediction-error models (Corlett *et al.*, 2010; Corlett and Fletcher, 2015). Similarly, two-factor theories often propose distinct mechanisms for different delusional misidentifications (Hirstein and Ramachandran, 1997; Ramachandran, 1998; Coltheart, 2007, 2010). However, we found a similar pattern of connectivity across cases of hyper and hypo familiarity and as well as misidentifications involving different categories of objects (e.g. people versus places). This commonality may explain why different types of delusional misidentifications can co-occur in the same patient (Paillère-Martinot *et al.*, 1994; Darby and Prasad, 2016). Determining whether connectivity differences exist between different types of delusional misidentifications will require a larger cohort of patients with non-overlapping symptoms.

Lesion network mapping

It is important to note that lesion network mapping does not involve obtaining functional neuroimaging from actual patients. Rather, functional connectivity from healthy normal subjects is used to determine regions normally connected to lesion locations. This approach has some advantages over

direct measurement of connectivity in actual patients (Boes *et al.*, 2015). First, this technique can be applied in rare syndromes such as lesion-induced delusional misidentifications, where imaging a patient cohort would be logistically difficult. Second, with patient data we would be unable to assess functional connectivity with a lesion location itself, as that region has been destroyed and has no neurophysiological activity. Finally, differences in functional connectivity observed in patients likely represent both lesion-induced dysfunction and compensatory responses. Lesion network mapping may identify regions more likely to be dysfunctional as a consequence of the lesion itself.

As with previous studies using lesion network mapping (Boes *et al.*, 2015), we found that both positive and negative correlation with the lesion location identified brain regions implicated in symptom expression. However, unlike this initial study, here we found that both positive and negative correlations identified regions likely involved in symptom expression in the same syndrome. Why one functional deficit would be based on positive correlation to the lesion location and another based on negative correlation remains unclear. One straightforward hypothesis is that regions positively correlated with a lesion location will have diminished activity, while anti-correlated regions will have increased activity, following the lesion. This simple framework is supported by analysis of lesions causing visual or auditory hallucinations, which are anti-correlated with regions that become hyperactive in these conditions (Boes *et al.*, 2015). However, there are also exceptions to this rule, such as post-stroke pain (Boes *et al.*, 2015). As such, we suggest that regions that are positively correlated or anti-correlated with a lesion location are likely to become dysregulated following the lesion, but whether this dysregulation is different depending on the sign of the correlation remains unclear.

Limitations

Several potential confounds in our lesion network mapping method have been previously addressed (Boes *et al.*, 2015). For example, a normative connectome can be used to approximate connectivity in patients (Fox *et al.*, 2014), results are similar with a connectome from older subjects age-matched to the lesion cohort (Boes *et al.*, 2015), and results are similar with different connectome processing strategies, including global signal regression (Boes *et al.*, 2015). We have also previously shown that 2D lesions from the literature can be used to approximate a 3D lesion (Boes *et al.*, 2015). However, we also confirmed this result in the present dataset using our two cases where 3D lesions were available (spatial correlations between the connectivity maps from 2D and 3D lesions were 0.91 and 0.89, respectively; see Supplementary Fig. 3).

Nevertheless, important limitations remain. First, lesion-network analysis identifies regions functionally connected to lesion sites, but this does not prove that these regions are dysfunctional in patients following the lesion. We

cannot currently predict whether regions connected to lesion locations will show increased activity, decreased activity, or a more complex pattern of dysfunction (Boes *et al.*, 2015). Functional neuroimaging in patients with active symptoms is necessary. This is difficult in rare syndromes; however, our prediction of retrosplenial dysfunction is supported by a functional MRI study of a single patient with lesion-induced Capgras (Thiel *et al.*, 2014). Our prediction of right frontal cortex dysfunction in these patients remains a testable but unproven hypothesis.

A second limitation is that tasks designed to measure expectation violation experimentally are likely an oversimplification of the ‘belief evaluation’ process thought to be necessary for rejection or perpetuation of a perceptual misidentification. Further, one could argue that selecting this task for our meta-analysis is a source of bias. These concerns are mitigated, but not eliminated by two analyses. First, we showed that our results remain significant using two additional meta-analyses of expectation violation: (i) an independent and pre-existing ALE meta-analysis of ‘oddball’ tasks; and (ii) a meta-analysis of expectation violation of cognitive beliefs. Second, we replicated connectivity to our expectation violation network in an independent cohort of lesion-induced delusion patients.

Conclusion

Because of the complex and bizarre nature of neuropsychiatric syndromes, it is often presumed that concurrent generalized cognitive dysfunction from delirium, psychiatric disease, dementia, or a second lesion is required (Levine and Grek, 1984; Rabins *et al.*, 1991; Devine *et al.*, 2014). Here we propose a mechanism by which a single lesion can cause complex symptoms based on that lesion’s precise location within the human connectome, without the need for pre-existing or hidden pathology. Our results are consistent with prior theories regarding delusional misidentifications (Ellis and Young, 1990; Hirstein and Ramachandran, 1997; Hirstein, 2010; Feinberg, 2011; Darby and Caplan, 2016) and theories of delusion formation more generally (Coltheart, 2007, 2010; Corlett *et al.*, 2010), and provide potential neuroanatomical correlates for impaired familiarity perception and belief evaluation in these patients.

Acknowledgement

We would like to thank Hongkeun Kim for sharing data for the oddball meta-analysis.

Funding

This work was supported by funding from the Sidney R. Baer, Jr. Foundation (R.D., S.L., M.F., A.P.L.), the NIH

(R01HD069776, R01NS073601, R21 NS082870, R21 MH099196, R21 NS085491, R21 HD07616 to A.P.L.; R25NS065743, K23NS083741 to M.F.), the Football Players Health Study at Harvard University (A.P.L.), and Harvard Catalyst | The Harvard Clinical and Translational Science Center (NCRR and the NCATS NIH, UL1 RR025758 to A.P.L.), and the American Brain Foundation (M.F.).

Supplementary material

Supplementary material is available at *Brain* online.

References

- Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 1994; 50: 7–15.
- Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. *Science* 1997; 275: 1293–5.
- Boes AD, Prasad S, Liu H, Liu Q, Pascual-Leone A, Caviness VS, et al. Network localization of neurological symptoms from focal brain lesions. *Brain* 2015; 138: 3061–75.
- Brighetti G, Bonifacci P, Borlimi R, Ottaviani C. “Far from the heart far from the eye”: evidence from the Capgras delusion. *Cogn Neuropsychiatry* 2007; 12: 189–97.
- Capgras J, Reboul-Lachaux J. Illusion des sosies dans un delire systematise chronique. *Bull la Soc Clin Med Ment* 1923; 2: 6–16.
- Coltheart M. Cognitive neuropsychiatry and delusional belief. *Q J Exp Psychol* 2007; 60: 1041–62.
- Coltheart M. The neuropsychology of delusions. *Ann N Y Acad Sci* 2010; 1191: 16–26.
- Corbetta M, Patel G, Shulman GL. The reorienting system of the human brain: from environment to theory of mind. *Neuron* 2008; 58: 306–24.
- Corlett PR, Aitken MRF, Dickinson A, Shanks DR, Honey GD, Honey RAE, et al. Prediction error during retrospective reevaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron* 2004; 44: 877–88.
- Corlett PR, Fletcher PC. Delusions and prediction error: clarifying the roles of behavioural and brain responses. *Cogn Neuropsychiatry* 2015; 20: 95–105.
- Corlett PR, Taylor JR, Wang X-J, Fletcher PC, Krystal JH. Toward a neurobiology of delusions. *Prog Neurobiol* 2010; 92: 345–69.
- Courbon P, Fail G. Syndrome d’illusion de Frégoli et schizophrénie. *Bull la Société Clin Médecine Ment* 1927; 15: 121–5.
- Darby R, Prasad S. Lesion-related delusional misidentification syndrome: a comprehensive review of reported cases. *J Neuropsychiatry Clin Neurosci* 2016; 28: 217–22.
- Darby RR, Caplan D. ‘Cat-gras’ delusion: a unique misidentification syndrome and a novel explanation. *Neurocase* 2016; 22: 1–6.
- Devine MJ, Bentley P, Jones B, Hotton G, Greenwood RJ, Jenkins IH, et al. The role of the right inferior frontal gyrus in the pathogenesis of post-stroke psychosis. *J Neurol* 2014; 261: 600–3.
- Devinsky O. Delusional misidentifications and duplications: right brain lesions, left brain delusions. *Neurology* 2009; 72: 80–7.
- Ellis HD, Young AW, Quayle AH, De Pauw KW. Reduced autonomic responses to faces in Capgras delusion. *Proc Biol Sci* 1997; 264: 1085–92.
- Ellis HD, Young AW. Accounting for delusional misidentifications. *Br J Psychiatry* 1990; 157: 239–48.
- Feinberg TE. *The lost self : pathologies of the brain and identity*. New York, NY: Oxford University Press, USA; 2005.
- Feinberg TE. Neuropathologies of the self: clinical and anatomical features *Conscious Cogn* 2011; 20: 75–81.
- Fox MD, Buckner RL, Liu H, Chakravarty MM, Lozano AM, Pascual-Leone A. Resting-state networks link invasive and noninvasive brain stimulation across diverse psychiatric and neurological diseases. *Proc Natl Acad Sci USA* 2014; 111: E4367–75.
- Geng JJ, Vossel S. Re-evaluating the role of TPJ in attentional control: contextual updating? *Neurosci Biobehav Rev* 2013; 37: 2608–20.
- Gobbini MI, Haxby JV. Neural systems for recognition of familiar faces. *Neuropsychologia* 2007; 45: 32–41.
- Goel V, Dolan RJ. Explaining modulation of reasoning by belief. *Cognition* 2003; 87: B11–22.
- Hirstein W, Ramachandran VS. Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons. *Proc Biol Sci* 1997; 264: 437–44.
- Hirstein W. The misidentification syndromes as mindreading disorders. *Cogn Neuropsychiatry* 2010; 15: 233–60.
- Kim H. Involvement of the dorsal and ventral attention networks in oddball stimulus processing: a meta-analysis. *Hum Brain Mapp* 2014; 35: 2265–84.
- Kumral E, Oztürk O. Delusional state following acute stroke. *Neurology* 2004; 62: 110–3.
- Levine DN, Grek A. The anatomic basis of delusions after right cerebral infarction. *Neurology* 1984; 34: 577–82.
- Metzack PD, Lavigne KM, Woodward TS. Functional brain networks involved in reality monitoring. *Neuropsychologia* 2015; 75: 50–60.
- Mitchell KJ, Johnson MK. Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? *Psychol Bull* 2009; 135: 638–77.
- Monakow C. *Die Lokalisation im Grosshirn : und der Abbau der Funktion durch kortikale Herde*. Wiesbaden: Verlag von J. F. Bergmann; 1914.
- Moretto G, Ládavas E, Mattioli F, di Pellegrino G. A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *J Cogn Neurosci* 2010; 22: 1888–99.
- Paillère-Martinot ML, Dao-Castellana MH, Masure MC, Pillon B, Martinot JL. Delusional misidentification: a clinical, neuropsychological and brain imaging case study. *Psychopathology* 1994; 27: 200–10.
- Pick A. On reduplicative paramnesia. *Brain* 1903; 26: 242–67.
- Posner MI. Orienting of attention. *Q J Exp Psychol* 1980; 32: 3–25.
- Rabins P V, Starkstein SE, Robinson RG. Risk factors for developing atypical (schizophreniform) psychosis following stroke. *J Neuropsychiatry Clin Neurosci* 1991; 3: 6–9.
- Ramachandran VS. Consciousness and body image: lessons from phantom limbs, Capgras syndrome and pain asymbolia. *Philos Trans R Soc Lond B Biol Sci* 1998; 353: 1851–9.
- Staton RD, Brumback RA, Wilson H. Reduplicative paramnesia: a disconnection syndrome of memory. *Cortex* 1982; 18: 23–35.
- Thiel CM, Studte S, Hildebrandt H, Huster R, Weerden R. When a loved one feels unfamiliar: a case study on the neural basis of Capgras delusion. *Cortex* 2014; 52: 75–85.
- Tranel D, Damasio H, Damasio AR. Double dissociation between overt and covert face recognition. *J Cogn Neurosci* 1995; 7: 425–32.
- Tranel D. Electrodermal activity in cognitive neuroscience: neuroanatomical and neuropsychological correlates. In: *Cognitive neuroscience of emotion*. New York, NY: Oxford University Press; 2000. p. 192–224.
- Vann SD, Aggleton JP, Maguire EA. What does the retrosplenial cortex do? *Nat Rev Neurosci* 2009; 10: 792–802.
- Young G. Capgras delusion: an interactionist model. *Conscious Cogn* 2008; 17: 863–76.
- Young G. In what sense ‘familiar’? Examining experiential differences within pathologies of facial recognition. *Conscious Cogn* 2009; 18: 628–38.